

AH 2004 : 3rd international conference on adaptive hypermedia and adaptive web-based systems : workshop proceedings part 1

Citation for published version (APA):

Aroyo, L. M., & Tasso, C. (Eds.) (2004). *AH 2004 : 3rd international conference on adaptive hypermedia and adaptive web-based systems : workshop proceedings part 1*. (Computer science reports; Vol. 0418). Technische Universiteit Eindhoven.

Document status and date:

Published: 01/01/2004

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Technische Universiteit Eindhoven
Department of Mathematics and Computer Science

AH 2004: 3rd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems

Workshops Proceedings
Part I

Editors and Co-Chairs:
Lora Aroyo
Carlo Tasso

04/18

ISSN 0926-4515

All rights reserved

Series editors: prof.dr. P.M.E. De Bra
 prof.dr.ir. J.J. van Wijk

Reports are available at:

<http://library.tue.nl/catalog/TUEPublication.csp?Language=dut&Type=ComputerScienceReports&Sort=Author&level=1> and

<http://library.tue.nl/catalog/TUEPublication.csp?Language=dut&Type=ComputerScienceReports&Sort=Year&Level=1>

Computer Science Reports 04/18
Eindhoven, 23rd August 2004

Table of Content

Preface

WORKSHOP 1:	
EAW'04: Engineering the Adaptive Web	1
Reasoning about Learning Object Metadata for Adapting SCORM Courseware <i>Matteo Baldoni, Cristina Baroglio, Viviana Patti and Laura Torasso</i>	4
Development of Software Tools for Designing Intelligent Tutoring Systems <i>Ildar Galeev, Larissa Tararina and Oleg Kolosov</i>	14
Adaptive Tourist Suggestions for a Commercial Web Site <i>Cristina Gena and Maria Orfino</i>	20
Personalization Functionality for the Semantic Web: Architectural Outline and First Sample Implementations <i>Nicola Henze and Matthias Kriesell</i>	26
Adaptive Web Information Systems: Architecture and Methodology for Reusing Content <i>Sébastien Iksal and Serge Garlatti</i>	36
Effectively Deploying an Adaptive Web Based eLearning System by Extending UML <i>Emma Quinn, Michael O'hEigearthaigh and Pramod Pathak</i>	46
Navigational Properties and User Attributes for Modelling Adaptive Web Applications <i>Gonzalo Rojas, Vicente Pelechano and Joan Fons</i>	52
Plasticity in Mobile Devices: a Dichotomic and Semantic View <i>Montserrat Sendin and Jesús Lorés</i>	58
Adaptive Hypermedia in the AdaptWeb Environment <i>Mariusza Warpechowski, José Palazzo Moreira de Oliveira, Maria Aparecida Martins Souto, Daniela Leal Musa, Maria Angélica de O. Camargo Brunetto, Veronice de Freitas, Mário Lemes Proença Jr.</i>	68
WORKSHOP 2:	
Individual Differences in Adaptive Hypermedia	74
Accommodating Learning Style Characteristics in Adaptive Educational Hypermedia Systems <i>Kyparissia A. Papanikolaou and Maria Grigoriadou</i>	77
End-User Quality of Experience Layer for Adaptive Hypermedia Systems <i>Cristina Hava Muntean and Jennifer McManis</i>	87
Evaluating Presentation Strategy and Choice in Adaptive Multiple Intelligence-based Tutoring System <i>Declan Kelly and Brendan Tangney</i>	97

Considering Human Memory Aspects to Adapting in Educational Hypermedia <i>Peter Agh and Maria Bielikova</i>	107
Discovery of Individual User Navigation Styles <i>Eelco Herder and Ion Juvina</i>	116
Designing an Adaptive Feedback Scheme to Support Reflection in Concept Mapping <i>Evangelia Gouli, Agoritsa Gogoulou, Kyparisia Papanikolaou and Maria Grigoriadou</i>	126
Adaptation Rules Relating Learning Styles Research and Learning Objects Metadata <i>Charalampos Karagiannidis and Demetrios Sampson</i>	136
Adapting Hypermedia to Cognitive Styles: Is it necessary? <i>Timothy Mitchell, Sherry Y. Chen and Robert Macredie</i>	146
WORKSHOP 3:	
Empirical Evaluation of Adaptive Systems	156
Proposed Evaluation Framework for Adaptive Hypermedia Systems <i>Arpita Gupta and P. S. Grover</i>	158
The First Click is the Deepest: Assessing Information Scent Predictions for a Personalized Search Engine <i>Karen Church, Mark T. Keane and Barry Smyth</i>	169
Evaluating Intelligent Tutoring Systems with Learning Curves <i>Brent Martin and Antonija Mitrovic</i>	179
Evaluation of Effects on Retrieval Performance for an Adaptive User Model <i>Hien Nguyen, Eugene Santos Jr., Qunhua Zhao and Chester Lee</i>	189
An Example of Evaluation Applied to a Course Adapted to Learning Styles of CHAEA's Test <i>M^a del Puerto Paule Ruiz, Juan Ramón Pérez Pérez and Martín González Rodríguez</i>	199
Evaluating a General-Purpose Adaptive Hypertext System <i>Chris Staff</i>	206
Evaluation Experiments and Experience from the Perspective of Interactive Information Retrieval <i>Ross Wilkinson and Mingfang Wu</i>	216

Preface

This volume contains the supplementary proceedings of the workshops organized in conjunction with AH'04, the third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, hosted by the Eindhoven University of Technology, August 23-26, Eindhoven, The Netherlands. The workshops proceedings consist of two parts:

Part I: Engineering and Evaluation of Adaptive Hypermedia Systems

- Workshop 1: Engineering the Adaptive Web
- Workshop 2: Individual Differences in Adaptive Hypermedia
- Workshop 3: Empirical Evaluation of Adaptive Systems

Part II: Adaptive Hypermedia Applications

- Workshop 4: Personalization in Future TV
- Workshop 5: Semantic Web for E-Learning
- Workshop 6: Authoring Adaptive and Adaptable Educational Hypermedia
- Workshop 7: Applying Adaptive Hypermedia Techniques to Service Oriented Environments: Pervasive Web Services and Context Aware Computing

While the main conference program presents an overview of the latest mature work in the field, the workshops at AH'04 aim at providing a wide international forum for researchers to present, discuss and explore their new ideas, work in progress and project developments within several focused areas of interest and within the overall scope of adaptive hypermedia systems. The workshops are intended to provide an informal interactive setting for participants to address in small focussed groups current technical and research issues related to the Adaptive Hypermedia and Adaptive Web-based Systems. Some of the core topics addressed in the AH'04 edition of workshops are:

- Emerging design methodologies for adaptive hypermedia (AH)
- Engineering of Adaptive Web-based Information Systems (AWIS)
- Design patterns for educational AH
- Authoring patterns and tools for educational AH
- Evaluation of AH authoring tools
- Adaptation and Semantic Web
- Semantic web-based educational AH architectures
- Web services for educational AH
- Existing metadata standards and ontologies for AH applications (e.g. digital TV, e-Learning, web services, individual differences, user modeling)
- Multi modalities and dimensions of individual differences in AHS
- Empirical criteria and methods AHS evaluation
- Personalized digital TV, T-commerce and T-learning

All workshop papers have been reviewed by committees of leading international researchers. We would like to thank each of the workshop organizers, including the program committees and additional reviewers, for their work in managing the review process and in the preparation and organization of their workshops.

August, 2004
Lora Aroyo and Carlo Tasso

WORKSHOP 1:
EAW'04: Engineering the Adaptive Web

**Methods & Technologies for Personalization and
Adaptation in the Semantic Web**

Workshop Co-Chairs:

Nicola Henze
Geert-Jan Houben

Preface

Fundamental research for building the Adaptive Web can be found in the area of Adaptive Hypermedia (AH): Instead of delivering all users the same kind of information ("one size fits all"), AH systems optimize the access to hypermedia-like systems according to the current, individual requirements of the particular user. The vision of the Semantic Web and the various standards and techniques developed within its context inspire a number of research areas (e.g. knowledge management, agent technology, information retrieval, adaptive hypermedia, web engineering) to target and facilitate more openness, interoperability and sharability of both resources and components. The application of those enabling technologies in the AH field leads to a new generation of AH that provides a basis for the 'Adaptive Web' - an adaptive information space enabling the sharing of metadata, resources and adaptation. It steps on an open component-based framework allowing efficient interoperability between various adaptive components and systems. Its goal is to enhance the existing AH applications in their adaptation to the user context, goals, preferences, or other environment specifics, and to allow for an optimal and personalized access to and navigation in distributed digital information - in so called Adaptive Web Information Systems (AWIS). The following questions focus on the core aspects for achieving the Adaptive Web:

- In the past, developing a Web application typically meant ad-hoc programming and a mixture of content and presentation design. The current data-intensive nature of (adaptive) Web Information Systems requires a more rigorous engineering and development process. In the light of constructing the user adaptation in an Adaptive Web Information System (AWIS) it is even more obvious that there are requirements for the engineering process. Design methodologies providing guidelines for the creation of modern AWIS are becoming essential. These methodologies should facilitate different aspects of AWIS design such as conceptual design, hypermedia design, presentation design, adaptation design etc. True Web engineering approaches offer designers and programmers a design framework often based on a model-driven approach, specifying the different aspects of the complete application design in terms of different models following the separation of concerns principle. Which design methodologies are emerging? How to facilitate engineering of AWIS?
- Machine understandable meaning can be expressed and conveyed with the help of metadata and ontologies. There are already a number of metadata specification standards as well as generally accepted subject domain and functional ontologies. Which technologies can facilitate the efficient use of existing metadata standards and ontologies for the purposes of AH applications?
- Various techniques are presently developed and explored within the context of Semantic Web or inspired by it (e.g. software agent communication, knowledge and information management strategies, metadata-based information retrieval, ontology and metadata-based information exchange, components sharing, open corpus hypermedia techniques). How can adaptation be implemented on top of these techniques in the context of adaptive web? What can support the sharing of adaptation functionality?

The aim of this workshop is to bring together researchers and practitioners in the fields of adaptive hypermedia, semantic web technologies, web engineering, software agents, knowledge management, information retrieval, user modeling, and other related disciplines which provide enabling technologies for personalization and adaptation on the World Wide Web. We hope that the workshop will explore how we can employ those enabling technologies within the adaptive hypermedia research as a significant step towards realizing the vision of the Adaptive Web.

August, 2004
Nicola Henze and Geert-Jan Houben

Program Committee

Lora Aroyo, *TU Eindhoven, the Netherlands*
Peter Brusilovsky, *University of Pittsburgh, USA*
François Bry, *Universität München, Germany*
Cristina Cachero, *Universidad de Alicante, Spain*
Owen Conlan, *Trinity College, Dublin, Ireland*
Rosa M. Carro, *Universidad Autonoma de Madrid, Spain*
Nicola Henze, *Universität Hannover, Germany (co-chair)*
Geert-Jan Houben, *TU Eindhoven, the Netherlands (co-chair)*
Eva Millan, *Universidad de Malaga, Spain*
Viviana Patti, *Universita di Torino, Italy*
Simos Retalis, *University of Cyprus, Cyprus*
Marcus Specht, *Fraunhofer-FIT, Germany*
Marlene Villanova-Oliver, *LSR-IMAG, France*
Howard Williams, *University of Edinburgh, UK*

Additional reviewers and local organizers

Peter Barna, *TU Eindhoven, the Netherlands*
Flavius Frasincar, *TU Eindhoven, the Netherlands*
Richard Vdovjak, *TU Eindhoven, the Netherlands*

Reasoning about learning object metadata for adapting SCORM courseware

Matteo Baldoni¹, Cristina Baroglio¹, Viviana Patti¹, and Laura Torasso^{1,2}

¹ Dipartimento di Informatica — Università degli Studi di Torino
C.so Svizzera, 185, Torino (Italy)

E-mail: {baldoni,baroglio,patti,torasso}@di.unito.it

² Società Reale Mutua di Assicurazioni — Servizio Informatico-S.I.Ge.A.
Via Corte d'Appello 11, Torino (Italy)

Abstract. In this work the problem of selecting and composing learning resources in the Semantic Web is considered. The starting point is the SCORM framework, used for the representation of learning objects. A proposal is done for describing a learning resource at the knowledge level, in terms of prerequisites and knowledge supplied, in order to enable the use of automated reasoning techniques (like planning) thus achieving forms of adaptation taken from the field of adaptive educational hypermedia. The description of learning strategies at the knowledge level opens the way to Semantic Web scenarios where learning resources are distributed over the network and reasoning systems can automatically select and compose them on-the-fly according to the user's needs. The advantages are an increase of reuse of the resources and a greater openness.

1 Introduction

The Semantic Web [6] is concerned with adding a semantic level to resources that are accessible over the internet in order to enable sophisticated forms of use and reuse. Resources are not all of a same kind; the most classical type of resource is the HTML document; recently, the attention has been posed also on software that can be invoked over the internet, leading to the definition of web services. Different proposals have been made for adding a semantic layer to the description of these resources, producing languages such as DAML+OIL and OWL for documents, OWL-S for web services. Especially with the development of peer-2-peer e-learning architectures [13], also *learning objects* can be considered as resources that are accessible over the internet, a view that is supported by some authors who report similarities between them and web services [2].

In the literature, there already exist various proposals for standardizing the description of learning objects, for instance to make them cross-platform (cross-LMS, learning management system). One of the most interesting is SCORM, especially in its new version 1.3 [14], which allows to describe a learning activity by including rules that govern the presentation of the learning items, by which the activity is composed, in an XML-based format.

The concept of “learning activity”, in a more general sense, draws considerably from the new teaching models proposed by pedagogy and psychology, in which a special attention is posed on the learner, once a passive listener and now a *promoter* of his/her own studies. Useless to say that the diffusion of the Internet greatly influenced this new perspective because, while in the traditional teaching style, the teacher was responsible of scheduling the lessons and of distributing the learning materials accordingly, the Web enabled the learner to have an “explorative” approach, in which he/she is free to focus on the preferred topics, to search for the learning objects across the world, and to choose the desired reading sequences. In order for navigation to be fruitful and personalized at the same time, however, the learner is to be supported in the exploration, for instance by taking into account his/her expertise when proposing new readings, or by forcing him/her to focus on some yet unknown elementary topic before passing to the study of an advanced feature.

In this framework it would be interesting to arrive to an integrated representation that, on a hand, takes into account the proposals of the standardization committees that work on learning object representation, while on the other it also takes into account the Semantic Web approach. In this way, it would be possible to apply the reasoning techniques that have been (and are being) developed in the Semantic Web area [1] to the problem of automatically selecting (over the internet) and composing learning objects, by adapting to the user’s learning goals and characteristics. In particular, we will show how techniques, that we have already applied to curriculum sequencing, can naturally be applied to this aim, given a proper extension of SCORM representations.

2 Background: AH and SCORM

In the last few years the field of adaptive hypermedia, applied to educational issues, attracted greater and greater attention [8]. Considerable advancements have been yield in the area, with the development of a great number of Web-based systems, like ELM-Art [15], the KBS hyperbook system [11], TANGOW [9], and many others, based on different, adaptive and intelligent technologies, with the common goal of using knowledge about the *domain*, about the *student* and about the *learning strategies* in order to support flexible, personalized learning and tutoring.

Among the technologies used in Web-based education for supporting adaptation and guidance, *curriculum sequencing*, where an “optimal reading sequence” through a hyper-space of learning objects is to be found, is one of the most popular [15, 11, 4]. Different methods have been proposed on how to determine which reading (or study) path to select or to generate in order to support in the best possible way the learner navigation through the hyper-space. However, following the definitions given in [3], it is useful to keep separate the *knowledge entities* or *competences*³ (i.e. some identifiable piece of knowledge related to

³ In this work we consider the two terms as synonyms.

the learning objects) and the *information entities* (that is the actual learning objects). Given such separation, it is possible to define at the *knowledge level*, a set of learning dependencies, that is the dependencies among knowledge entities (or competences). We can, then, associate to each learning object a set of competences that describe it. In this framework, it is possible to add to the system an adaptation component, that uses such a knowledge, together with a representation of the user *learning goal* and of the user knowledge, for performing the sequencing task, producing sequences that fit the user requirements and characteristics, based on the available learning objects.

Working at the level of competences is closer to human intuition and makes the reuse of the learning objects easier because the same learning object will be automatically taken into account by the adaptation component whenever a competence that is supplied by it is necessary during the sequencing process. Moreover, it enables the application of *goal-directed reasoning processes*, as it is done by the WLog system [4]. In this system the learning objects are represented as actions each having a set of preconditions (competences that are necessary for using the learning object) and a set of effects (the supplied competences). Competences can be connected by causal relationships. A group of agents, called *reasoners*, uses such descriptions, the user learning goal (expressed as well in terms of competences) for performing the sequencing task. This is done by refining curriculum schemas, described only on the basis of the defined knowledge entities, and decoupled from the actual learning objects. Thus, adaptation is based on the *reasoning capabilities* of the *rational agents*, that are implemented in the logic language DyLOG [5]. The reasoning techniques that are used by the agents are taken from the field of “reasoning about actions” and are *planning*, *temporal projection*, and *temporal explanation*; basically, they allow reasoning about the dynamics of the learning objects outcomes and preconditions and to generate sequences of learning objects for achieving the learning goal.

On the other hand, talking about learning objects representation, there is a need for a standardized framework which not only describes them but it also rules their presentation. SCORM is one such framework, which is attracting greater and greater attention, and is supported both by commercial and by open source platforms. In SCORM 1.3 terminology the learning units are called SCO, and their structure plus the rules, that govern the learning activity, are defined in the so-called “manifest” of the SCO. Broadly speaking each manifest describes both the structure into which the learning material is assembled and the way in which it is presented. The language by which rules are written basically exploits three operators: sequencing, if-then branching, and presentation of a set of learning items that the user can freely explore. These operators allow the description of a learning object as a tree in which inner nodes (items) represent sub-activities. The tree leaves are the single units (assets) of which the learning object is made (e.g. a set of HTML pages). The decision by which the next item to show is taken by the Learning Management System (LMS), based on the rules contained in the manifest and on features that depend on the user behavior (e.g. the user has read the previous item, the user has not answered a question correctly). The

nice point is the intrinsic modularity of this representation: learning objects can be composed, they can be reused in many compositions, and reuse can occur at any level, so composed learning objects can be reused as well as a whole.

Each SCO can be annotated by adding a description in terms of IEEE LOM (Learning Object Metadata). More specifically, a complete LOM description [10] consists of attributes, divided in nine categories (general, life cycle, meta-metadata, technical, educational, rights, relation, classification, and annotation). In [13] it is shown how fifteen of such attributes are sufficient to describe most of the learning resources. Such attributes include the possibility of describing the *contents* of a learning object in terms of keywords taken from an *ontology* of interest. Therefore, in principle, by means of LOM it is possible to include in a SCO a description at the level of knowledge entities (we will come back to this point); it would, then, be possible to apply reasoning techniques, of the kind described shortly above: it would be possible to dynamically assemble the learning objects to be used in a course, on the basis of the learning goals, to verify if a learning object satisfies a given learning goal, or to adapt a general learning strategy to a user's needs. To this aim, the architecture of the Learning Management System

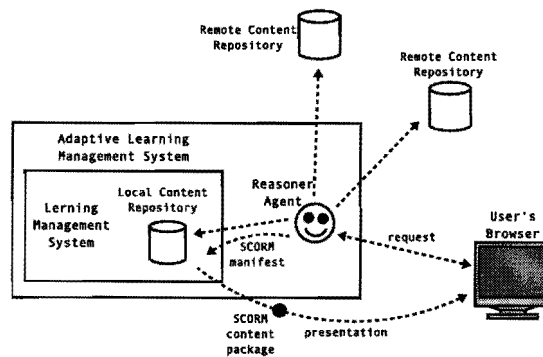


Fig. 1. Architecture of a Learning Management System augmented with a reasoning component.

could be extended by introducing a new, “intelligent” component (see Figure 1) which, on a side, interacts with the user (or with a requester agent) for collecting the desired learning goals and goal conditions, while on the other it can query the local and external repositories for selecting proper learning objects, that it will, in some cases, also assemble.

3 Adding a knowledge level to SCORM learning objects

Following what done in [4], we can interpret a learning object as an action: an action can be executed given that a set of conditions holds, by executing it, a set

of conditions will become true. According to this metaphor, a learning object can profitably be used if the learner has a given set of prerequisite competences; by using it, the learner will acquire a new set of competences. So, the idea is to introduce at the level of the learning objects, some metadata that describe both their *pre-requisites* and *effects*, as done in the curriculum sequencing application.

Regarding annotation, LOM allows the annotation of the learning objects by means of an ontology of interest (see for instance [13]), by using the attribute *classification*. A LOM classification consists of a set of ontology elements (or *taxons*), with an associated role (the *purpose*). Figure 2 shows an example. The taxons in the example are taken from the DAML version of the ACM computer classification system ontology [12]. The reference to the ontology is contained in the *source* element. Since the XML-based representation is quite long, for the sake of brevity only two taxons have been reported: the first (relational database) is necessary in order to understand the contents of the learning object, while the other (scientific databases) is a competence that is supplied by the learning object.

The proposed annotation expresses a set of *learning dependencies* in terms of *knowledge entities*. Such learning dependencies can be expressed in a declarative formalism, and can be used by a reasoning system. Given a set of learning objects, annotated by pre-requisites and effects, it is possible to compose reading sequences by using the standard planners, that have been developed by the Artificial Intelligence community, for instance, the well-known Graphplan (first described in [7]). Graphplan is a general-purpose planner that works in STRIPS-like domains; as all planners, the task that it executes is to build a sequence of atomic actions, that allows the transition from an initial state to a state of interest, or goal state. The algorithm is based on ideas used in graph algorithms: it builds a structure called *planning graph*, whose main property is that the information that is useful for constraining the plan search is quickly propagated through the graph as it is built.

General-purpose planners search a sequence of interest in the whole space of possible solutions and allow the construction of learning objects on the basis of any learning goal. However, this is not always adequate in an educational application framework, where the set of learning goals of interest, in that context, is fairly limited and the experience of the teachers, in structuring the courses and the learning materials, is important. For instance, a teacher, who has been assigned a new course, may express that a topic *A* is to be presented before topic *B*. This kind of constraint cannot be exploited by a general-purpose planner unless topic *A* is an effect of some learning object that supplies competences requested by *B* as preconditions. The organization of the learning materials not only depends on strict prerequisites but it is also up to the *experience* of the teacher, i.e. it is necessary to consider also the *view* of the teacher on how the learning object should be structured.

On the other hand, it is not reasonable to express schemas in terms of specific learning objects. The ideal solution is to express the afore-mentioned schemas as *learning strategies*, i.e. a rule (or a set of rules) that specifies the overall

```

<lom xmlns="http://www.imsglobal.org/xsd/imsmd_vip2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.imsglobal.org/xsd/imsmd_vip2 imsmd_vip2p2.xsd">
  <general>
    <title>
      <langstring>module A</langstring>
    </title>
  </general>
  ...
  <classification>
    <purpose>
      ...
      <value><langstring>Prerequisite</langstring></value>
    </purpose>
    <taxonpath>
      <source>
        <langstring>http://daml.umbc.edu/ontologies/classification.daml</langstring>
      </source>
      <taxon>
        <entry>
          <langstring xml:lang="en">relational database</langstring>
        </entry>
      </taxon>
    </taxonpath>
  </classification>
  ...
  <classification>
    <purpose>
      ...
      <value><langstring>Educational Objective</langstring></value>
    </purpose>
    <taxonpath>
      <source>
        <langstring>http://daml.umbc.edu/ontologies/classification.daml</langstring>
      </source>
      <taxon>
        <entry>
          <langstring xml:lang="en">scientific databases</langstring>
        </entry>
      </taxon>
    </taxonpath>
  </classification>
</lom>

```

Fig. 2. Excerpt from the annotation for the learning object 'module A': "relational database" is an example of prerequisite while "scientific databases" is an example of educational objective.

structure of the learning object, expressed only in terms of *competences*. The construction of a learning object can, then, be obtained by refining a learning strategy, according to specific requirements and, in particular, by choosing those SCOs, that are the most suitable to the student. As we will see in the next section, we propose to represent a learning strategy as a declarative program. Notice that all its possible executions satisfy the learning goals of the strategy. Adaptation, in this case, consists in selecting an execution that also satisfies the specific user's requirements.

4 Introducing learning strategies

Learning strategies, as well as learning objects, should be defined on the basis of an ontology of interest. Besides supplying a vocabulary of common terms, as it happens in many cases, ontologies also express *part-of* or *is-a* relations between the terms in the classification. So, for instance, in the already mentioned ACM ontology, *relational databases* is part of *database management*, as well as *query languages*, *distributed databases*, and *scientific databases*. In other words, the ontology says that if a resource is annotated by the word *relational databases*, then it explains something about *database management*; it does not say that in order for *database management* to be true *relational databases* must necessarily be true.

Learning strategies, however, can better be defined by exploiting other relations between the knowledge entities. One common need is to express *conjunctions* or *sequences* of *knowledge entities*. So for instance, one can say that in his/her view, it is possible to acquire competence about *database management* only by getting competence about *all* of its subclasses mentioned above, and that *relational databases* must be known before *distributed databases* is introduced.

An example that we consider particularly meaningful is preparing the material for a basic computer science course: the course may have different contents depending on the kind of student to whom it will be offered (e.g. a Biology student, rather than a Communication Sciences student, rather than a Computer Science student). Hereafter, we consider the case of Biology students and propose a DyLOG procedure, named '*strategy('informatics_for_biologists')*', that expresses, at high level, a learning strategy for guiding a biology student in a learning path, which includes the basic concepts about how a computer works, together with a specific competence about databases. Notice that no reference to specific learning objects is done.

```
strategy('informatics_for_biologists') is
  achieve_goal(has_competence('computer system organization')) ∧
  achieve_goal(has_competence('operating systems')) ∧
  achieve_goal(has_competence('database management')).
...
achieve_goal(has_competence('database management')) is
  achieve_goal(has_competence('relational databases')) ∧
  achieve_goal(has_competence('query languages')) ∧
```

$achieve_goal(has_competence('distributed\ databases')) \wedge$
 $achieve_goal(has_competence('scientific\ databases'))$.

strategy is defined as a procedure clause, that exploits the view of the strategy creator on what it means to acquire competence about *computer system organization*, *operating systems*, and *database management*. Observe that, for avoiding collision between the definition of a label in the ontology of reference, and the view that the strategy creator has on how that knowledge entity could be achieved, a renaming should occur. For the sake of simplicity, however, we have not renamed the labels used in the example.

For instance, supposing that the name of the SCORM learning object at issue is *module A*, we could represent in DyLOG its learning dependencies, originally written in LOM as described by Figure 2, in the following way:

$access(learning_object('module\ A'))$ **possible if**
 $has_competence('distributed\ database') \wedge$
 $has_competence('relational\ database')$.
 $access(learning_object('module\ A'))$ **causes**
 $has_competence('scientific\ databases')$.

In the case of DyLOG representations, given a learning strategy, it is possible to apply *procedural planning* for refining it and possibly assemble a new learning object made of SCOs, that are annotated with the competences, suggested by the strategy. Opposite to general-purpose planners, procedural planning searches for a solution in the set of executions of a learning strategy. Notice that, since the strategy is based on competences, rather than on specific resources, the system might need to select between different courses, annotated with the same desired competence, which could equally be selected in building the actual learning path. This choice can be done based on external information, such as a user model, or it may be derive from a further interaction with the user. All these steps should be carried on by the intelligent component added to the LMS architecture (see Figure 1). The resulting plan can be stored as a SCORM manifest, which can be considered as an instance of the original learning strategy. Decoupling the strategies from the learning objects results in a greater flexibility of the overall system, in a greater ease of reuse of the learning objects, and on the possible (partial) automatization of the construction of ad hoc learning objects. As well as learning objects, also learning strategies could be made public and shared across different systems.

5 Conclusions

In this paper we have discussed the advantages of applying curriculum sequencing techniques from the field of adaptive hypermedia to the problem of generating personalized SCORM-based courses that build on learning objects potentially distributed on the semantic web. The current technology already allows the annotation of learning objects in a way that enables the application of Semantic

Web concepts and techniques. In particular, it is possible to profit of the LOM *classification* attribute, for describing a learning resource at the knowledge level, in terms of prerequisite competences and competence supplied, where competences are entries of some shared ontology.

Such a kind of annotation supports the interpretation of a learning object, written according to the SCORM framework, as an action having precondition and effects, and then opens the way to the application of standard Artificial Intelligence reasoners for performing various tasks. In particular we focussed on building on-the-fly learning objects that allow the achievement of a learning goal of interest, based on already available learning material, making use of a representation of learning strategies in the high level logic programming language DyLOG. Our description of learning strategies is based on competences, rather than on specific resources, a fundamental key for opening the way to Semantic Web scenarios, where learning resources are distributed over the network and reasoning systems make use of semantic annotation for automatically selecting and composing them, according to the user's needs. The advantages are an increase of reuse of the resources and a greater openness. DyLOG supports procedural planning; given a learning strategy description, it allows to find a learning path through the learning material that fulfills both the user goals and the strategy guidelines. Procedural planning constrains the search space of solutions, a particularly relevant question when the number of available resources is big, as it might be on the web. Resulted solutions can be translated in SCORM manifests for the presentation to the user, thus we can interpret a SCORM manifest as an instance of a learning strategy, i.e. a presentation that respects the guidelines given by it, combining specific SCOs. Such an instance is adapted to the particular user goal. This level of adaptation is currently missing in the SCORM courseware generation module. In fact the kind of adaptation that is currently offered is very simple and it is based exclusively on the navigation behavior of the user. An item is shown if the user has already visited one or more other items or if he has given the wrong answer to a question associated to such an item. However, the structure of the course is given and cannot be built on the fly adapting to the user current goals. We can say that the two kind of adaptation are orthogonal: by reasoning we compose personalized learning paths; then, such learning paths are presented as manifests and the adaptation techniques based on monitoring the user behavior, already supported by the LMS, can be applied for achieving a further step of adaptation.

Acknowledgement

This research has been partially funded by the European Commission and by the Swiss Federal Office for Education and Science within the 6th Framework Programme project REVERSE number 506779 (cf. <http://reverse.net>).

References

1. S. Abdennadher, J. Alves Alferes, G. Antoniou, U. Assmann, R. Backofen, C. Baroglio, P. A. Bonatti, F. Bry, W. Drabent, N. Eisinger, N. E. Fuchs, T. Geisler, N. Henze, J. Maluszynski, M. Marchiori, A. Martelli, S. Carro Martinez, H. Jurgen Ohlbach, S. Schaffert, M. Schroeder, K. U. Schulz, U. Schwertel, and G. Wagner. Automated reasoning on the web. *Communications of Applied Logic*, 2004. To appear.
2. L. Aroyo, S. Pokraev, and R. Brussee. Preparing SCORM for the semantic web. In *Proc. of the Int. Conf. on Ontologies, Databases and Applications of Semantics (ODBASE03)*, Catania, Italy, 2003.
3. M. Baldoni, C. Baroglio, N. Henze, and V. Patti. Setting up a framework for comparing adaptive educational hypermedia: First steps and application on curriculum sequencing. In *Proc. of ABIS-Workshop 2002: Personalization for the mobile World, Workshop on Adaptivity and User Modeling in Iterative Software Systems*, pages 43–50, Hannover, Germany, October 2002.
4. M. Baldoni, C. Baroglio, and V. Patti. Web-based adaptive tutoring: an approach based on logic agents and reasoning about actions. *Artificial Intelligence Review*, 2004. To appear.
5. M. Baldoni, L. Giordano, A. Martelli, and V. Patti. Programming Rational Agents in a Modal Action Logic. *Annals of Mathematics and Artificial Intelligence, Special issue on Logic-Based Agent Implementation*, 41(2-4), 2004. To appear.
6. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, 2001.
7. A. Blum and M. Furst. Fast planning through planning graph analysis. *Artificial Intelligence*, 90:281–300, 1997.
8. P. Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11:87–110, 2001.
9. R.M. Carro, E. Pulido, and P. Rodriguez. Dynamic generation of adaptive internet-based courses. *Journal of Network and Computer Applications*, 22:249–257, 1999.
10. Learning Technology Standards Committee. Draft Standard for Learning Object Metadata, IEEE 1484.12.1-2002, 2002.
11. Nicola Henze and Wolfgang Nejdl. Adaptation in open corpus hypermedia. *IJAIED Special Issue on Adaptive and Intelligent Web-Based Systems*, 12:325–350, 2001.
12. Association of Computing Machinery. The ACM computer classification system, 2003. <http://www.acm.org/class/1998/>.
13. W. Nejdl P. Dolog, R. Gavriloaie and J. Brase. Integrating adaptive hypermedia techniques and open rdf-based environments. In *Proc. of The 12th Int. World Wide Web Conference*, Budapest, Hungary, 2003.
14. ADL Technical Team. SCORM XML controlling document - SCORM CAM version 1.3 navigation XML XSD version 1.0, 2004. <http://www.adlnet.org/>.
15. Gerhard Weber and Peter Brusilovsky. ELM-ART: An Adaptive Versatile System for Web-based Instruction. *IJAIED Special Issue on Adaptive and Intelligent Web-Based Systems*, 12(4):351–384, 2001.

Development of software tools for designing intelligent tutoring systems

Ildar Galeev, Larissa Tararina, Oleg Kolosov

Kazan State Technological University, Russia
monap@kstu.ru

Abstract. Software tools MONAP have been considered in the paper. The mentioned tools provide designing of a partially integrated learning environment, which includes an electronic textbook (ET) and an intelligent tutoring system (ITS). Adaptation to the learner is based on the skills overlay model. A new, additional adaptation method, integrated into software tools, has been described as well. . Relevant for the learner educational material, represented in ET, is determined at every training step on the basis of the tutorial problem solution results. The learner is offered to study first of all the pages, which were worse mastered.

1 Introduction

By the present time, a lot of methods and techniques of adaptive hypermedia have been described [1]. Adaptation in educational hypermedia is based on the level of the learner's knowledge. The knowledge state of particular learners, as a rule, changes in different ways during the teaching and learning process. To describe the learner's knowledge of the subject, an overlay model based on the subject domain model is most commonly used. The domain model is represented as a network of domain concepts. For every domain model concept the overlay model of each particular learner contains certain values, which are used to assess the level on which he knows this concept. Hence, the overlay model of the learner's knowledge is represented as a set of pairs "concept – value" for every domain concept and serves to implement adaptation technologies in educational hypermedia.

Investigations on adaptive educational technologies have been undertaken in the Department of Information Technologies and Teaching Aids of Kazan State Technological University (KSTU) over a period of several years [2-5]. In the recent works the authors have considered a number of problems of building partially integrated learning environments, where adaptation is based on the overlay model of the learner's skills, represented as a set of pairs "operation (rule) – value"[4, 5]. The overlay model is created and maintained in intelligent tutoring systems (ITS) designed with the help of software tools of the MONAP family [2, 3]. The overlay model of skills provides the basis for the adaptive control of the learner, consisting in giving him a tutorial problem of an optimal difficulty and complexity for the next training step. Integrated environment, including an electronic textbook (ET) and an

intelligent tutoring system, gives additional possibilities in raising the level of the teaching/learning process adaptability. Such environments require the development of new integrated methods of adaptation to the learner, providing, in particular, presentation of relevant educational material according to the results of solving the tutorial problem.

2. Student's knowledge identification

The kernel of the overlay model of the learner's skills is represented as the vector $P(k) = [P_1(k), P_2(k), \dots, P_j(k), \dots, P_J(k)]$, where $P_j(k)$ is the probability of using the j -type operation correctly. This probability is calculated with the application of Bayesian approach and the results of the tutorial problem at the k -th step of training [2].

The student's knowledge identification ($P_j(k)$ value determination) is carried out in the following manner. N hypotheses H_i ($i=1,2,\dots,N$) corresponding to N being-learned states are input for every y_j operation. For every i -th being-learned state there is a conditional probability $P(A_j/H_i)$ of correct y_j operation application in each of its L_j applications, equal to $\frac{i}{N+1}$.

The hypotheses H_i form a full group of antithetical events, i.e. $\sum_{i=1}^N P_{ij} = 1$ occurs, where P_{ij} is the probability of the hypotheses H_i for the operation y_j .

At every learning step the event $B_j(k)$ takes place that consists in correct application of j -th operation $M_j(k)$ times from $L_j(k)$ defined.

This information serves to re-calculate the distribution of the hypotheses P_{ij} probabilities using the Bayesian formula.

Every k -th learning step is characterized by a priori and a posterior distribution of probabilities of the hypotheses for being-trained states $P_{ij}^0(k)$ and $P_{ij}^1(k)$ connected by the following relationship:

$$P_{ij}^1(k) = \frac{P_{ij}^0(k) \times P(B_j(k)/H_i)}{\sum_{s=1}^N P_{sj}^0(k) \times P(B_j(k)/H_s)}, \quad (1)$$

where $P(B_j(k)/H_i)$ is determine by the Bernoulli theorem, i.e.

$$P(B_j(k)/H_i) = C_{L_j(k)}^{M_j(k)} \times P(A_j/H_i)^{M_j(k)} \times (1 - P(A_j/H_i))^{L_j(k)-M_j(k)}, \quad (2)$$

where $C_{L_j(k)}^{M_j(k)}$ is the number of $L_j(k)$ combinations $M_j(k)$.

Taking into consideration the fact that a priori distribution of the hypotheses probabilities at the k -th step coincides with a posteriori distribution at the $(k-1)$ -th step, i.e. $P_{ij}^0(k) = P_{ij}^1(k-1)$ takes place, formula (1) may be re-written in the form which emphasizes its recursive character (the whole history of learning is allowed for), namely:

$$P_{ij}^1(k) = \frac{P_{ij}^1(k-1) \times P(B_j(k)/H_i)}{\sum_{s=1}^N P_{sj}^1(k-1) \times P(B_j(k)/H_s)}. \quad (3)$$

The probability of the operation y_j correct application at the k -th step is determined with the formula of complete probability:

$$P(A_j(k)) = \sum_{i=1}^N P_{ij}^1(k) \times P(A_j/H_i). \quad (4)$$

The final evaluation $P_j(k)$ results from the reduction of the value calculated with formula (4) to the input being-learned states.

The mistake control and necessary explanation output at the k -th learning step provides a means for predicting the probability of the operation y_j correct application at the $(k+1)$ -th learning step:

$$P_j(k+1/k) = V \times P_j(k), \text{ where } V = \frac{P_j(k)}{P_j(k-1)}. \quad (5)$$

3. Determination of the relevant educational material

Establishing interconnections between operations and concepts provides presentation of relevant educational material to the learner according to the results of solving the tutorial problem. Interconnection of operations (rules) and concepts reflects the relation $E \subseteq Y \times X$, where Y is a multitude of operations, and X – a multitude of concepts. This relation is specified by the matrix $\|e_{ji}\|$, its lines corresponding to the operations $y_1, y_2, \dots, y_j, \dots, y_J$, and columns – to the concepts

$x_1, x_2, \dots, x_t, \dots, x_T$. The element of the matrix $\|e_{jt}\|$ is determined in the following way:

$$e_{jt} = \begin{cases} 1, & \text{if concept } x_t \text{ is used in operation } y_j; \\ 0, & \text{in the opposite case.} \end{cases}$$

Figure 1 shows the example of the matrix $\|e_{jt}\|$ with an additional lower line, containing elements $\alpha_t (t = \overline{1, T})$, where $\alpha_t = \sum_j e_{jt}$ is a number of operations from the whole multitude of operations Y , in which the concept x_t is used.

OPERATIONS (RULES)	CONCEPTS					
	x_1	x_2	...	x_t	...	x_T
y_1	1	1	...	0		0
y_2	0	1	...	0		1
...
y_j	1	0		1		0
...
y_J	0	0		1		1
	α_1	α_2	...	α_t	...	α_T

Fig. 1. Matrix of interconnection between operations and concepts

Matrix $\|e_{jt}\|$ is a formalized description of the operation structures. The concept used in the operation is its part and parcel. Hence, at every k -th step of learning the level of the learner's mastering of every concept used in the operation y_j cannot be lower than the level of mastering this operation as the whole at the same step, i.e. $\omega_t(k) \geq P_j(k)$, where $\omega_t(k)$ is the evaluation of the level of mastering the concept x_t . In the general case it is impossible to calculate $\omega_t(k)$ more properly by the result of performing the operation y_j , because there is no adequate feedback. As a rule, a qualifying dialogue with the learner provides the necessary feedback. The dialogue is subject-dependent and, thus, cannot be used to develop methods of evaluating the learner's knowledge invariant to SD. In the context of the fact, that when the tutorial problem is being solved at the k -th step of learning, one and the same concept can be used in different operations, in order to calculate the integral estimate of

the concept x_i mastering levels, it is necessary to take into consideration these operations mastering level. So, to calculate the integral estimate of the concept x_i mastering level $\omega_i(k)$ by the results of executing the tutorial task at the k -th training step, the following formula is proposed:

$$\omega_i(k) = \frac{\sum_j e_{ji} \cdot P_j(k)}{\alpha_i}. \quad (6)$$

In a similar manner, the interconnection between ET pages and the concepts, described in these pages, specifies the interconnections between operations and concepts (Fig.1). This interconnection is reflected by the relation $F \subseteq S \times X$, where S is the multitude of pages of ET, and X – the multitude of concepts. Relation F is specified by the matrix $\|f_{gi}\|$, its lines corresponding to pages $s_1, s_2, \dots, s_g, \dots, s_G$, and columns – to concepts $x_1, x_2, \dots, x_i, \dots, x_T$. The element of the matrix $\|f_{gi}\|$ is determined in the following way:

$$f_{gi} = \begin{cases} 1, & \text{if concept } x_i \text{ is described in page } s_g; \\ 0, & \text{in the opposite case.} \end{cases}$$

The estimate of the learner's non-mastering $\beta_g(k)$ the knowledge, given in the ET at page s_g , at the k -th training step is calculated according to the formula:

$$\beta_g(k) = \frac{\sum_i f_{gi} \cdot (1 - \omega_i(k))}{\sum_i f_{gi}}. \quad (7)$$

Using $\omega_i(k)$ values calculated on the basis of $P_j(k)$ values, when calculating $\beta_g(k)$, makes it possible to take into account the prehistory of the tutorial task execution by the learner, this being an important requirement at building adaptive tutorial systems.

When the task is finished but there are mistakes, a message is formed, containing a list of references to the ET pages and telling the learner to return to the work over the educational material given there. The list of ET pages is sorted according to $\beta_g(k)$ values decrease, i.e. the learner is offered to study first of all the pages, which were worse mastered.

The implementation of the proposed method of determining relevant ET pages, corresponding to the learner's knowledge and skills at the k -th training step, re-

quired to widen the learning environment, created and maintained by software tools of MONAP family [3,4].

To provide the link between the ET concepts and the ITS operations, a new parameter of the integrated learning environment – “the number of the concepts being studied” – has been introduced.

Together with the modification of the basis RULE, the basis CONCEPTS, being a matrix $\|f_{gr}\|$, is developed and maintained. A properly organized thesaurus, included into the ET general information components, can fulfill the function of the matrix $\|f_{gr}\|$ as well. In this case it is not necessary to create the basis CONCEPTS.

4. Conclusions

The efficiency of the computer-assisted teaching and learning process is mostly determined by the progress, achieved when solving the following problems: development of multifunctional integrated learning environments and provision of adaptive learning in these environments. In line with the purpose in hand, a partially integrated learning environment has been developed. This environment includes ET and ITS. Adaptation to the learner is based on the skills overlay model, designed and maintained by software tools of the MONAP family. A new, additional adaptation method, integrated into software tools, has been developed as well. Relevant for the learner educational material, represented in ET, is determined at every training step on the basis of the tutorial problem solution results. Invariance of the described adaptation mechanisms to SD makes them potentially easily replicated.

5. References

1. Brusilovsky, P. Methods And Techniques Of Adaptive Hypermedia. User Modeling and User-Adapted Interaction, 6 (2-3), 1996, pp. 87-129
2. Galeev, I. Automation of the ETS Desing, Educational Technology - September-October 1999. - V. XXXIX, No. 5. - pp. 11-15.
3. Galeev I., Chepegin V. and Sosnovsky S., MONAP: Models, Methods and Applications, Proceedings of the International Conference KBCS 2000, Mumbai, India. pp. 217-228.
4. Ildar Galeev, Larissa Tararina and Oleg Kolosov, “Problems of building adaptive integrated learning environments”, in Vladan Devedzic, J. Michael Spector, Demetrios G. Sampson and Kinshuk (ed): Proceedings of 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'2003), Athens, Greece, July 9-11, 2003, p. 486.
5. Ildar Galeev, Larissa Tararina, Oleg Kolosov, Vlad Kolosov, Structure and implementation of partially integrated adaptive learning environment, in Allison Rossett (ed): Proceedings of E-Learn 2003, Phoenix, Arizona USA, November 7-11, 2003, p. 2151-2154.

Adaptive tourist suggestions for a commercial web site

Cristina Gena¹, Maria Orfino²

¹Dipartimento di Informatica
Università degli Studi di Torino
Corso Svizzera 185, 10149 Torino, Italy

²Corso di Laurea in Scienze della Comunicazione - Università degli Studi di Torino
cgena@di.unito.it, m_orfino@virgilio.it

Abstract. This paper describes a commercial web site that offers tourist information about Turin and the valleys hosting the 2006 Olympic Winter Games. Since this site offers a large amount of information for foreign and non-resident users, we decided to offer adaptive suggestions in order to filter the items presented to the users and to give orientation about the localization of the presented attractions. In this initial adaptation phase we decided to generate the adaptations by applying commonsense rules related to the features of the presented items without asking any registration from the users.

1. Introduction

In recent years the popularity of the personalization concept is increasing and now many companies try to offer personalized services to their customers. Since in today competitive business environment providing value to the customers is of great importance for companies to survive, they try to satisfy users' needs and wishes, giving them a high-quality product/service at the right time in the right way. The same principles inspire the development of *personalized hypermedia applications* (also defined as *customer relationship applications* [6]). These systems adapt "content, structure, and/or presentation of the networked hypermedia objects (web pages) to each individual user's characteristics, usage behaviour and/or usage environment" [6]. In the tourist field, several personalized hypermedia applications have been developed. Citysearch [2] is a web site that offers tourist information about American cities. The system asks the user to choose the city in which the navigation starts and the specific area of that city. Then the user can select a specific category, like Hotel and a specific item she is interested in. At that point the system provides her with a list of related "Search nearby" links concerning restaurants, bars, museums, etc, she can find in the neighbourhood of that hotel. GUIDE [4] presents information to the users tailored to both environmental context (the major attractions in the city) and the visitor's personal context, for example her current location, her profile (composed by

¹ The authors are grateful to the WebWorking srl staff for the help in the development of the system and for their precious suggestions.

her interests), and the set of attractions already visited. WebGuide [5], which is intended for web-based and mobile tourist guides and aimed at providing personalized tour recommendations for the city of Heidelberg on the basis of geographical information, information about points of interest and selected means of transport, individual user's interests and preferences, and tour restrictions specified by the user. Intrigue [1] is a prototype tourist information server that presents information on desktop and handset devices. This system recommends sightseeing destinations and itineraries by taking into account the preferences of heterogeneous tourist groups and provides an interactive agenda for scheduling the tour and on the basis of user's choices and time constraints generates a suitable itinerary.

We believe that the exploitation of adaptivity in the tourist field could help foreign and non-resident people to better satisfy their needs by receiving tailored information on the basis of their interests and their preferences. The use of adaptive systems seems to be the right answer which supports information needs and tailors the information on the basis of the visitor's interests, giving her a useful tool for finding her bearings in the city.

In this paper we focus on the generation of adaptive suggestions in a commercial web site (<http://www.club2006.com>) that is aimed at being a source of up to date tourist information for those people who are going to visit Turin, its outskirts and the two alpine valleys hosting the 2006 Olympic Winter Games. The paper is organized as follows: Section 2 describes the web site and the adaptive rules we implemented in order to generate the adaptive suggestions presented on the web site. Section 3 concludes the paper and presents future directions.

2. Overview of the system and its adaptive features

The Club2006 web site is aimed at providing up to date information about restaurants, hotels, bars, attractions and events that will take place in Turin, in its outskirts and in the two alpine valleys (Susa and Pinerolese valleys) hosting the 2006 Olympic Winter Games. The site is structured in eight main sections: *things to do*, *where to stay*, *where to eat*, *how to get*, *living here*, *shopping*, *services*, *my Club2006*². In this initial implementation of adaptations, we decided to concentrate our work on the first three sections: *i) things to do*, which offers to the user information about events, art, history, sport, and about excursions organized in the towns and in the countryside; *ii) where to stay*, which contains information about hotels, farmhouses, bed & breakfast, self-catering and weekly hires; *iii) where to eat*, which presents restaurants, pizzerias, wine bars and cafés, farmhouses, and ethnic restaurants.

All these sections require the selection of the starting place (Turin, its outskirts, Susa Valley, Pinerolese Valleys) at the beginning of the interaction. For instance, *i)* the user clicks on *Where to stay*, *ii)* then she may select the place (Turin, its outskirts, Susa Valley, Pinerolese Valleys), *iii)* and finally she can choose the subcategory she is interested in (e.g., hotels); *iv)* at that point the system will present a list of items (e.g., a list of hotels), each one conducting to a more detailed page containing a brief

² This section is devoted to Club2006's partners that can personalize their presentation pages hosted in the Club2006 web site.

description (e.g., information about the selected hotel such as address, rooms, pictures, reservations, etc).

The site maintains all the information in several Microsoft Access® databases and then dynamically generates web pages by exploiting Active Server Pages (ASP) server-side scripting environment.

We decided to develop an adaptive version of Club2006 web site for its main characteristics: *i*) the databases contain a large amount of records for every section. Particularly, the more commercial sections contain vast lists of items since the main purpose of the site is to attract foreign visitors by offering a great choice of on-line hotel reservations and detailed information about the main attractions. Therefore, the user can benefit from the intelligent filter in the presentation of the required information; *ii*) since the target users of Club2006 are foreign visitors or Italian people coming from other parts of Italy, they can benefit from a virtual compass to orientate themselves and explore in advance the town they are about to visit. This kind of information is nowadays offered by on-line applications accessible from mobile devices. However, web sites are still the main source of information because, at the moment, the exploitation of handheld devices during free time is not so popular and in addition, not so many services are available [8]. Therefore, we decided to add recommendations by offering adaptive orientation information in order to help users in a virtual tour of the towns presented in the Club2006 web site.

As described above, we initially added adaptive features to three sections of the site. We decided not to require from users to register. This last point can be an advantage, under certain points of view, since it is well known that adaptable sites require additional user effort that can cause the paradox of the active user [3]. Users often refuse to visit the sites that require to complete a questionnaire because it takes time that would get their immediate task done. Also, Manber et al. [7] observed that the majority of users do not customise the web pages when they have the possibility to do it. They prefer a quick answer to their questions that only a tailored solution can give. Unfortunately, most of adaptive sites require a period of interaction before they are able to show the advantages of personalised pages to the users.

In order to benefit from the adaptive information without exploiting any information about specific users, we applied commonsense rules, not tailored towards a specific user but adapted to the features of the item selected. The basic idea was to offer adaptive suggestions useful to the user's orientation. So, for instance, when the user selects a specific item (e.g., a hotel) belonging to one of the three categories previously described, the system generates a short and non-obtrusive list of suggestions of nearby attractions complementary to the selected item (e.g., restaurants, bars, museums).

In order to further adapt the information to the user's need, we decided to apply another filter by considering the economic range in which some of the items presented to the user can be classified. We exploited the large Club2006 partners' database and we applied the "1 to 5 stars" classification technique, typical of hotels, in order to classify the other attractions on the basis of their price level. The basic idea behind this reasoning is that people typically choose where to eat and where to stay on the basis of their budget. So, for instance, it is more common that young tourists choose hostels and bed and breakfast, cheap pizzerias and fast-foods, while older and well-off people tend to stay in higher stars hotel and to eat in more expensive restaurants.

Therefore, our adaptation mechanism considers not only the geographical area to produce recommendations, but also the stars classification applied to all the commercial records of the database. For the recommendations concerning non-commercial items (such as museums, palaces, monuments) we decided to select the more representative items as points of reference in every quarter of the town. For instance, the famous Egyptian Museum of Turin will be classified as point of reference in that quarter (and therefore suggested when a user selects an attraction in that quarter) instead of a minor art gallery where an exhibition takes place.

Both for commercial and non-commercial categories, we re-classified the items adding meta-information in every record we considered. For commercial records we added stars by applying an automatic mechanism, while for each non-commercial record we manually selected the points of reference.

The adaptive suggestions are generated by applying adaptation rules based on the following formalism:

If a user chooses x^{l-s} (1 to 5 stars item) in the zip code z then suggests $(y_1^{l-s*}, y_2^{l-s*}, y_3, y_4^{l-s*})$ in the zip code z*

where the y_n^{l-s*} star items have the same stars as x^{l-s*} item selected by the user and the *zip code* defines precise areas in which the city is divided

For instance, considering the category *Where to stay*, the x (user's selection) and y variables (suggestions of the system) can assume the following values

x = where to stay

y_1 = where to eat (subcategory: restaurants and pizzerias)

y_2 = where to eat (subcategory: wine bar and café)

y_3 = where to go (a point of references defined for that area)

For every commercial variable y (suggestions belonging to *where to eat* and *where to stay* categories) only three items are presented to the user by choosing between the most requested items: at the top of the list there will be presented the items that have received the large number of clicks. For non-commercial items y (*where to go*) the more representative point of reference of the selected area is presented. The recommendations are shown on the left-hand side of the page (see Figure 1) and each item is a link leading to a presentation page containing a textual description, images and other specific information about the selected item.

For instance, in Figures 1 the following rule is applied:

IF the user chooses Jolly Hotel Ligure**** in the 10122 zip code

THEN suggests

- "Il Carignano", "Il Biricchino", "Albero di vino" (**** stars restaurants in 10122 zip code)
- "Ozio Caf ", "Noh Cool Caf ", "Petit Fleur" (**** stars caf s in 10122 zip code)
- "The Egyptian Museum" (point of reference of 10122 zip code)



Fig. 1. Adaptive suggestions on the left side of the page.

3. Conclusions and future works

The adaptation rules implemented in the presented system are lightweight, intuitive and they can be easily ignored from who does not believe in personalization advantages or does not want a computer deciding for herself. The system does not require to log in, so the user does not waste time filling in registration forms and she is not afraid of leaving her personal data for commercial purposes. In addition, the generation of the adaptive suggestions does not add complexity to the system since the required meta-information exploited by the adaptation rules are stored in the same database that the system queries in order to extract the regular information.

However, these recommendations are merely based on commonsense rules and are not tailored towards interests of a specific user. Therefore they can be less effective than more tailored suggestions based on reasoning concerning implicit and explicit information about the users. At the moment we are implementing the rules on the on line web site together with a monitoring mechanism aimed at checking the real exploitation of the adaptive suggestions by online users. We decided on the evaluation of real users interacting with the system in practice by analyzing their real behaviour. Therefore, we will be able to evaluate suggestions and correlations by monitoring actions such as: the user clicks on the suggestion, the user prints or saves the suggested information, etc. Moreover, we are analyzing log files in order to discover associations and regularities in the path of real users. Another future work will concern the generation of more tailored suggestions for those users who voluntarily register into the system in order to obtain hotel reservations, subscribe to the mailing list, participate to the forum, etc.

The adaptive suggestions offered by Club2006 web site are quite similar to the Citysearch's "nearby" information, as described in Section 2. Both sites propose suggestions about attractions that lie in the neighbourhood of a specific place, but Club2006 also considers the economic classification (1 to 5 stars) of suggested items and orders them on the basis of their popularity. In addition, we added contextualized information such as the points of reference in order to help the orientation in a foreign town. Other differences between the two sites concern the amount of suggestion categories taken into account and the selection of the area in which the navigation starts. Citysearch allows the users to receive "nearby" information about restaurants, hotels, bars, department stores, shopping centers and malls, theatres, banks, florists, etc, while Club2006, gives information merely about places where to eat and where to stay and references about the major cultural attractions. In future, we want to implement other personalization rules in order to suggest additional items related to other categories, such as shopping and events. Concerning the selection of the initial area, Citysearch maintains the filter and presents localized information until the user selects another area of the city. However, this behaviour can sometimes disorient the user because during the first interactions it can be difficult to comprehend the functioning of the system and to learn how to obtain information without the area filter or to change the filter. Moreover, Club2006 does not require such a detailed selection of the areas, since at the moment the towns presented on the site are not so widespread and so well-known to foreign visitors.

References

1. L. Ardissono, A. Goy, G. Petrone, M. Segnan, P. Torasso, "INTRIGUE: personalized recommendation of tourist attractions for desktop and handset devices", in *Applied Artificial Intelligence, Special Issue on Artificial Intelligence for Cultural Heritage and Digital Libraries*, vol.17, n.8-9, Taylor & Francis, 2003, 687-714.
2. Citysearch, <http://www.citysearch.com>
3. J.M Carroll and M.B. Rosson, "The Paradox of the Active User", in J.M. Carrol (Ed.), *Interfacing Thought: Cognitive Aspects of Human-Computer Interaction*, Cambridge, MA: MIT Press, 1987, pp.80-11.
4. K. Cheverst, K. Mitchell, N. Davies, "The Role of Adaptive Hypermedia in a Context-Aware Tourist Guide", in *Communications of ACM* 45, May 2002, pp.47-51.
5. Fink, J. and A. Kobsa, "User Modeling in Personalized City Tours", *Artificial Intelligence Review*, 18(1), 2002, 33-74.
6. A. Kobsa, J. Koenemann, W. Pohl: *Personalized Hypermedia Presentation Techniques For Improving Online Customer Relationship*, in *The Knowledge Engineering Review*, Vol. 16(2), Cambridge University Press, 2001, p.111-155
7. U. Manber, A. Patel and J. Robison, "Experience with personalization of Yahoo!", *Communication of ACM* 43, 8, Aug.2000, 35 - 39.
8. R., Scalisi, "L'Affermazione del Mobile Computing. Indagine sugli Utenti dei Computer Palmari", PhD thesis, Università degli Studi di Torino, 2003.

Personalization Functionality for the Semantic Web: Architectural Outline and First Sample Implementations

Nicola Henze¹ and Matthias Kriesell²

¹ ISI – Semantic Web Group, University of Hannover,
Appelstr. 4, D-30167 Hannover, Germany
`henze@kbs.uni-hannover.de`

² Institut für Mathematik A, Welfengarten 1
D-30167 Hannover, Germany
`kriesell@math.uni-hannover.de`

Abstract. We propose a service-based architecture for bringing methods and techniques from the area of adaptive hypermedia to the Semantic Web. In our framework, personalization functionalities from adaptive hypermedia are available as web-services which a user can subscribe / un-subscribe as s/he likes. We have implemented our ideas in the *Personal Reader*, a framework for defining rule-based personalization algorithms for Semantic Web applications. In this paper, we present the basic architecture of the Personal Reader framework, and describe its realization in one example instance: A Personal Reader for displaying learning resources for Java programming. Other instances like e. g. a Personal Reader for publications, are currently under development.

Keywords: adaptive hypermedia, personalization, adaptive web, semantic web, reasoning on the semantic web, rules for the semantic web.

1 Introduction

With the idea of a Semantic Web [1] in which machines can understand, process and reason about resources to provide better and more comfortable support for humans in interacting with the World Wide Web, the question of personalizing the interaction with web content is at hand.

In the area of adaptive hypermedia, research has been carried out to understand how personalization and adaptation strategies can be successfully applied in hypertext systems and hypertext like environments. It has been stated that in the area of adaptive hypermedia and of adaptive web-based systems, the focus of developed systems has been so far on *closed world* settings. This means that these systems work on a fixed set of resources which are normally known to the system designers at design time (see the discussion on closed corpus adaptive hypermedia [4]). This observation also relates to the fact that the issue of authoring adaptive hypermedia systems is still one of the most important research questions in this area, see e. g. [2]

A generalization of adaptive hypermedia to an *Adaptive Web* [6] depends therefore on a solution of the closed corpus problem in adaptive hypermedia. In this paper, we propose an architecture for applying *some* of the techniques developed in adaptive hypermedia to an open corpus. In the Personal Reader project³ we are developing a framework for designing, implementing and maintaining web content readers, which provide personalized enrichment of web content for the individual user. We will describe the basic architecture and its realization (section 2). As an example reader, a “Personal Reader for Learning Resources” is described in section 3. At the end of the paper, we compare our approach to related work in this area and conclude with an outlook on current and future work.

2 Architecture

The question on how to enable personalization functionality in the Semantic Web can be regarded from different viewpoints, involving different disciplines, e. g. data mining, machine learning, web graph analysis, collaborative approaches, adaptive hypermedia. In our approach, we concentrate on methods and techniques developed in the area of adaptive hypermedia. An analysis and comparison framework, based on a logical description of adaptive (educational) hypermedia systems, has been presented in [14]. Here, some typical methods used in adaptive hypermedia systems, have been described as rules in first order logic. Required data (metadata about the documents, the users, as well as runtime data like observations about user interactions, etc.) have been identified and described. The approach presented in this paper is based on this catalogue of adaptive functionality and discusses an implementation hereof for the Semantic Web.

The architectural outline for implementing the Personal Reader is a rigorous approach for applying Semantic Web technologies. A modular framework of components / services - for visualizing the Personal Reader and providing the user interface, for mediating between user requests and available personalization services, for user modeling, for providing personal recommendations and context information, et cetera, is the basis for the Personal Reader. The communications between all components / services is syntactically based on RDF descriptions. E.g. the request for getting personal recommendations for a learning resource for a certain user is provided by an RDF description which is exchanged between the components mediator and personal recommendations. Thus a component is a services, which is usually independent from the others and which can interact with them by “understanding” the RDF notifications they send (see figure 1). The common “understanding” is realized by referring to semantics in the ontologies used in the RDF descriptions which provide the valid vocabulary.

In the following we will present the main ideas on how RDF enables the communication, how learning resources and domain concepts are annotated for

³ www.personal-reader.de

the Personal Reader, and which ontologies or standards are required and used for the Personal Reader.

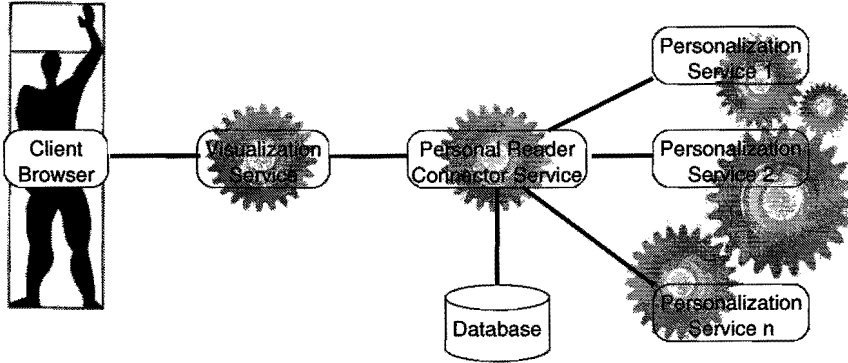


Fig. 1. Architecture of the Personal Reader framework, showing the different components of the Personal Reader: Visualization (user interface), the Personal Reader backbone (consisting of the Connector Services, the Reasoning Service(s)), and some data-provision services, for RDF data and for the connection with some database for storing user profile information.

2.1 Ontologies for Describing the Objects of Discourse

As a basic implementation paradigm, we decided to describe all *objects of discourse* in our framework using RDF / RDF-Schema (Resource Description Framework and -Schema, [19]) or a higher-level ontology language like the Web Ontology Language (OWL) [18]. In particular, we employ the following ontologies for describing our objects of discourse:

1. a domain ontology describing the application domain, and a document ontology. We assume that documents are annotated according to standard metadata schemas for documents like e.g. Dublin Core (DC) [11], or, in the area of education, according to the Learning Objects Metadata standard (LOM) [17];
2. a user model ontology (attribute-value pairs for user characteristics, preferences, information on the devices the user is using for accessing the Personal Reader, etc.);
3. an observation ontology (for describing the different kinds of user observations made during runtime);
4. and an adaptation ontology for describing the adaptation functionality which is provided by the adaptation services.

It is important to note that we refer in the Personal Reader framework as far as possible to standard metadata annotations: E. g. in the sample reader we present in this paper, the metadata descriptions of documents are in accordance with LOM, user profile information is relying on the IEEE PAPI specification for describing learners [15]. Further, we apply domain ontologies, in the example a domain ontology for Java programming. By using ontologies for describing run-time user observations and for adaptation, these models can be shared with other applications, however, there are currently no standards for these kinds of ontologies available. Due to space constraints, we will not elaborate on the ontologies further in this paper; more details can be found e.g. in [13].

2.2 Reasoning

Each personal learning service possess reasoning rules for some specific adaptation purposes. These rules query for resources and metadata, and reason over distributed data and metadata descriptions. A major step for reasoning after having queried the user profile, the domain ontology, and learning objects is to construct a temporally valid task knowledge base as a base for applying the adaptation rules. We will present some examples of adaptation rules in section 3 where we present a Personal Reader for learning resources.

For implementing the reasoning rules, we decided to use the TRIPLE query and rule language for the Semantic Web [20]. Rules defined in TRIPLE can reason about RDF-annotated information resources (required translation tools from RDF to triple and vice versa are provided). An RDF statement (which is a triple) is written as `subject[predicate -> object]`

RDF *models* are explicitly available in TRIPLE: Statements that are true in a specific model are written as "`@model`". This is particularly important for constructing the *temporal knowledge bases* as required in the Personal Reader. Connectives and quantifiers for building logical formulae from statements are allowed as usual: AND, OR, NOT, FORALL, EXISTS, <-, ->, etc. are used.

2.3 Administration

The administration component of the Personal Reader framework allows us to easily integrate new instances of Readers. E. g. in the e-learning domain, the integration of course materials which are — at least — described to our standard, and for which some domain ontology exists, can immediately be integrated and displayed in the Personal Reader. The flexibility of the Triple language, especially the support of models, allows us in the Personal Reader framework to realize personalization functionality in accordance to course descriptions *and* domain ontologies, or to course descriptions alone.

3 Example: Personal Reader for the Sun Java Tutorial

In this section, we present an example of a Personal Reader instance: A Personal Reader for learning resources. We have implemented the Reader for displaying

the learning resources of the Sun Java Tutorial [8], a freely available online Tutorial on Java programming.

This Personal Reader helps the learner to view the learning resources in a context: In this context, more *details* related to the topics of the learning resource, the *general topics* the learner is currently studying, *examples*, *summaries*, *quizzes*, etc. are generated and enriched with personal recommendations according to the learner's current learning state, as shown in figure 2.

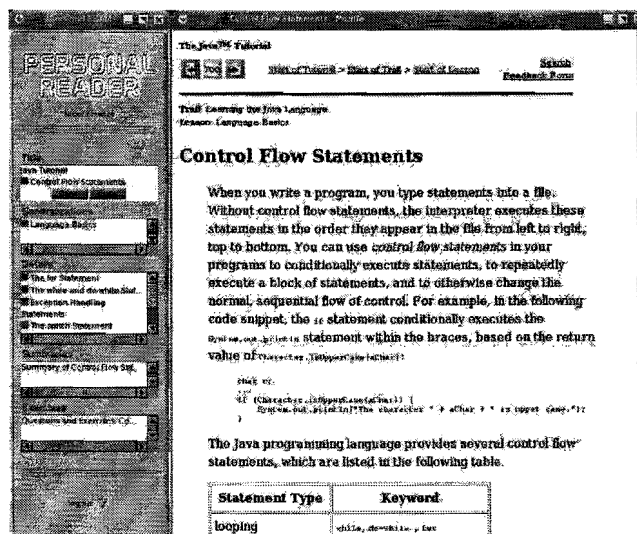


Fig. 2. Screenshot of the Personal Reader, showing the adaptive context of a learning resource in a course.

In this section we discuss how we implemented the adaptation rules for the adaptive context generation.

3.1 Reasoning in a Personal Reader for Learning Resources

In the following, we will describe some of the rules that are used by the Personal Reader for learning resources to determine appropriate adaptation strategies.

Providing a context by displaying details of a learning resource Generating links to more detailed learning resources is an adaptive functionality in this example Personal Reader.

The adaptation rule takes the isA hierarchy in the domain ontology, in this case the domain ontology for Java programming, into account to determine domain concepts which are details of the current concept or concepts that the learner is studying on the learning resource. In particular, more details for the

currently used learning resource is determined by `detail_learningobject(LO, LO_DETAIL)` where `LO` and `LO_Detail` are learning resources, and where `LO_DETAIL` covers more specialized learning concepts which are determined with help of the domain ontology.

```
FORALL LO, LO_DETAIL detail_learningobject(LO, LO_DETAIL) <-
  EXISTS C, C_DETAIL(detail_concepts(C, C_DETAIL)
    AND concepts_of_LO(LO, C) AND concepts_of_LO(LO_DETAIL, C_DETAIL))
    AND learning_resource(LO_DETAIL) AND NOT unify(LO,LO_DETAIL).
```

N. B. the rule does neither require that `LO_DETAIL` covers all specialized learning concepts, nor that it exclusively covers specialized learning concepts. Further refinements of this adaptation rule are of course possible and should, in a future version of the Personal Reader, be available as tuning parameters under control of the learner. The rules for embedding a learning resource into more general aspects with respect to the current learning progress are similar.

Providing pointers to Quizzes Another example of an *adaptation rule* for generating embedding context is the recommendation of quiz pages. A learning resource `Q` is recommended as a quiz for a currently learned learning resource `LO` if it is a quiz (the rule for determining this is not displayed) and if it provides questions to at least some of the concepts learned on `LO`.

```
FORALL Q quiz(Q) <-
  Q['http://www.w3.org/1999/02/22-rdf-syntax-ns#':type ->
    'http://ltsc.ieee.org/2002/09/lom-educational#':Quiz']
FORALL Q, C concepts_of_Quiz(Q,C) <-
  quiz(Q) AND concept(C) AND
  Q['http://purl.org/dc/elements/1.1/':subject -> C].
FORALL LO, Q quiz(LO, Q) <-
  EXISTS C (concepts_of_LO(LO,C) AND concepts_of_Quiz(Q,C)).
```

Calculating Recommendations. Recommendations are personalized according to the current learning progress of the user, e. g. with respect to the current set of course materials. The following rule determines that a learning resource `LO` is recommended if the learner studied at least one more general learning resource (`UpperLevelLO`):

```
FORALL LO1, LO2 upperlevel(LO1,LO2) <-
  LO1['http://purl.org/dc/terms#':isPartOf -> LO2].

FORALL LO, U learning_state(LO, U, recommended) <-
  EXISTS UpperLevelLO (upperlevel(LO, UpperLevelLO) AND
    p_obs(UpperLevelLO, U, Learned) ).
```

Additional rules deriving stronger recommendations (e. g., if the user has studied *all* general learning resources), less strong recommendations (e.g., if one or two of these haven't been studied so far), etc., are possible, too.

Recommendations can also be calculated with respect to the current domain ontology. This is necessary if a user is regarding course materials from different courses at the same time.

```
FORALL C, C_DETAIL detail_concepts(C, C_DETAIL) <-
  C_DETAIL['http://www.w3.org/2000/01/rdf-schema#':subClassOf -> C]
  AND concept(C) AND concept(C_DETAIL).

FORALL LO, U learning_state(LO, U, recommended) <-
  EXISTS C, C_DETAIL (concepts_of_LO(LO, C_DETAIL)
    AND detail_concepts(C, C_DETAIL) AND p_obs(C, U, Learned) ).
```

However, the first recommendation rule, which reasons within one course will be more accurate because it has more fine-grained information about the course and therefore on the learning process of a learner taking part in this course. Thus, our strategy is to apply first the adaptation rule which take most observations and data into account, and, if these rules cannot provide results, apply less strong rules. In future work, we will extend this approach. Currently, we are considering in enriching the results of the rules with confidence parameters. How these confidence values can be smoothly integrated into a user interface is an open research question.

Reasoning Rules for User Modeling The Personal Reader requires only view information about the user's characteristics. Thus, for our example we employed a very simple user model: This user model traces the users path in the learning environment and registers whenever the user has visited some learning resource. This information is stored in the user's profile, which is binded to RDF as follows:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:j.0="http://semweb.kbs.uni-hannover.de/rdf/l3s.rdf#" >
  <rdf:Description rdf:about="http://semweb.kbs.uni-hannover.de/user#john">
    <rdf:type rdf:resource="http://hoersaal.../rdf/l3s.rdf#User"/>
    <j.0:hasVisited>http://java.sun.com/.../variables.html</j.0:hasVisited>
    ...
```

From this information, we derive whether a particular user learned some concept. The following rule derives all learned concepts.

```
FORALL C, U p_obs(C, U, Learned) <-
  EXISTS LO (concepts_of_LO(LO, C) AND
    U['http://semweb.kbs.uni-hannover.de/rdf/l3s#':hasVisited ->LO]).
```

Similarly, it can be determine whether a learning object has been learned by a user.

4 Related Work

Related work to our approach includes standard models of adaptive hypermedia like [2], recent personalization systems [12, 9] as well as personalized learning portals [7].

Comparing our work with standard models for adaptive hypermedia systems like e.g. AHAM [3], we observe that they use several models like conceptual, navigational, adaptational, teacher and learner models. Compared to our approach, these models either correspond to ontologies / taxonomies, to different schemas describing teacher and learner profile, and to schemas describing the navigational structure of a course. We express adaptation functionalities as encapsulated and reusable Triple rules, while the adaptation model in AHA uses a rule based language encoded into XML. AHA! provides the strategies for adaptation at the resources [2]. [12] focuses on content adaptation, or, more precisely, on personalizing the presentation of hypermedia content to the user. The technique used here is a slice-technique, inspired by the Relationship Management Methodology [16]. Both adaptability and adaptivity are realized via slices: Adaptability is provided by certain adaptability conditions in the slices, e. g., the ability of a device to display images. Adaptivity is based on the AHAM idea [3] of event-conditions for resources: A slice is desirable if its appearance condition evaluates to true.

[10] builds on separating learning resources from sequencing logic and additional models for adaptivity: Adaptivity blocks in the metadata of learning objects as well as in the narrative model, candidate groups and components define which kind of adaptivity can be realized on the current learning content. Driving force in these models are the candidate groups that define how to teach a certain learning concept. A rule engine selects the best candidates for each user in a given context. Adaptivity requirements are considered only in the adaptivity blocks. Personalized learning portals are investigated in [7]. The learning portals provide views on learning activities which are provided by so-called *activity servers*. The activity servers store both learning content and the learning activities possible with this special content. A central student model server collects the data about student performance from each activity server the student is working on, as well as from every portal the student is registered to. In [5], also *value-added services* are introduced in the architecture. The architecture in our approach is a simplification of the architecture presented here: We only consider value-added services, and implemented our personalization services as these value-added services.

5 Conclusion

We have presented a framework for designing, implementing and maintaining adaptive *reader* applications for the Semantic Web. The Personal Reader framework is based on the idea of establishing personalization functionality as services on the (Semantic) Web. The realization of personalization functionality is done

on the logic layer of the Semantic Web tower, making use of description and rule language recently developed in the context of the Semantic Web. We have tested the framework with an example reader, the Personal Reader for the Sun Java programming tutorial. Currently, we are using the framework to design a Reader for publications, and are investigating how learner assessment can be integrated to enhance the functionality for learning resources. The current state of the project can be followed at www.personal-reader.de.

6 Acknowledgment

This work has been partially supported by the European Network of Excellence REVERSE - Reasoning on the Web with Rules and Semantics⁴. We would like to thank also Stefan Decker, Peter Dolog, Andreas Harth, Marc Herrlich, Wolfgang Nejdl, and Michael Sintek for their support in getting the idea of a Personal Reader into reality.

References

1. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.
2. P. D. Bra, A. Aerts, D. Smits, and N. Stash. AHA! version 2.0: More adaptation flexibility for authors. In *Proceedings of the AACE ELearn'2002 conference*, pages 240–246, Oct. 2002.
3. P. D. Bra, G.-J. Houben, and H. Wu. AHAM: A dexter-based reference model for adaptive hypermedia. In *ACM Conference on Hypertext and Hypermedia*, pages 147–156, Darmstadt, Germany, 1999.
4. P. Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11:87–110, 2001.
5. P. Brusilovsky. KnowledgeTree: A distributed architecture for adaptive e-learning. In *Proceedings of the Thirteenth International World Wide Web Conference*, New York, May 2004.
6. P. Brusilovsky and M. Maybury. *The Adaptive Web*. Communications of the ACM, 2002.
7. P. Brusilovsky and H. Nijhawan. A framework for adaptive e-learning based on distributed re-usable learning activities. In *In Proceedings of World Conference on E-Learning, E-Learn 2002*, Montreal, Canada, 2002.
8. M. Campione and K. Walrath. The java tutorial, 2003. <http://java.sun.com/docs/books/tutorial/>.
9. O. Conlan, D. Lewis, S. Higel, D. O'Sullivan, and V. Wade. Applying adaptive hypermedia techniques to semantic web service composition. In *International Workshop on Adaptive Hypermedia and Adaptive Web-based Systems (AH 2003)*, Budapest, Hungary, 2003.
10. D. Dagger, V. Wade, and O. Conlan. Towards “anytime, anywhere” learning: The role and realization of dynamic terminal personalization in adaptive elearning. In *Ed-Media 2003, World Conference on Educational Multimedia, Hypermedia and Telecommunications*, Hawaii, 2003.

⁴ www.reverse.net

11. Dublin Core, 2004. <http://dublincore.org/>.
12. F. Frasincar and G. Houben. Hypermedia presentation adaptation on the semantic web. In *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2002)*, Malaga, Spain, 2002.
13. N. Henze, P. Dolog, and W. Nejdl. Reasoning and ontologies for personalized e-learning. *ETS Journal Special Issue on Ontologies and Semantic Web for eLearning*, 2004. To appear.
14. N. Henze and W. Nejdl. Logically characterizing adaptive educational hypermedia systems. In *International Workshop on Adaptive Hypermedia and Adaptive Web-based Systems (AH 2003)*, Budapest, Hungary, 2003.
15. IEEE P1484.2 Learner Model Working Group, PAPI Learner, draft 7 specification, 2003. http://ltsc.ieee.org/wg2/papi_learner_07_main.pdf.
16. T. Isakowitz, E. A. Stohr, and P. Balasubrahmanian. RMM: A methodology for structured hypermedia design. *Commun. ACM*, 38(8), August 1995.
17. LOM: Draft Standard for Learning Object Metadata, 2002. <http://ltsc.ieee.org/wg12/index.html>.
18. OWL, Web Ontology Language, W3C Recommendation, Feb. 2004.
19. Resource Description Framework (RDF) Schema Specification 1.0, 2002. <http://www.w3.org/TR/rdf-schema>.
20. M. Sintek and S. Decker. TRIPLE - an RDF Query, Inference, and Transformation Language. In I. Horrocks and J. Hendler, editors, *International Semantic Web Conference (ISWC)*, pages 364–378, Sardinia, Italy, 2002. LNCS 2342.

Adaptive Web Information Systems: Architecture and Methodology for Reusing Content

Sébastien Iksal¹, Serge Garlatti²

¹LIUM – University of Maine – IUT de Laval
52 rue des Docteurs Calmette et Guérin
BP 2045 – 53020 LAVAL Cedex 09, France
+33 2 43 59 49 26
sebastien.iksal@univ-lemans.fr

²IASC Dept. – GET – ENST Bretagne
Technopôle de Brest Iroise
CS 83818, 29238 BREST Cedex, France
+33 2 29 00 14 53
serge.garlatti@enst-bretagne.fr

Abstract. Nowadays, adaptive web information systems use partially the Web to provide different kinds of content, navigation tools and layouts according to user needs. We focus on AWIS for which users share a common knowledge to work together. For us, AWIS design is an intensive knowledge driven process. We propose the methodology and architecture used in the flexible composition engine called SCARCE. The paper presents the key issues for reusing the content in the methodology: interoperability and W3C standards, consistency of the delivered document and the distinct specification and management of AWIS components. The main benefits of this approach are: i) a generic AWIS architecture which is reusable in different contexts, ii) this architecture is tuned to the explicit knowledge of communities and provide a method for AWIS design.

Keywords. Adaptive Web Information Systems, Composition Engine, Semantic Web, Metadata, User Model.

1 Introduction

Adaptive web information systems – AWIS - have the ability to provide different kinds of content, navigation tools and layouts according to user needs [1]. Traditional AWIS used to have a domain model, a user model and adaptation rules. The domain model is used to convey resource semantics. The content is typically known to the author of the application and under his control [2]. Thus, the website structure, its content, its adaptation strategies and its presentation are often designed by the same author and are often combined.

Nowadays, AWIS use partially the Web as an information space where resources are distributed. They must have the ability to reuse distributed data repositories [2, 3] and/or web services [4]. One can view AWIS consisting of different components: an organization, content, adaptation strategies and presentation. We focus our study on a methodological approach enhancing the content reuse. Some key issues of that are: interoperability and W3C standards, consistency of the delivered document and the distinct specification and management of AWIS components. The current information-intensive nature of web information systems requires more rigorous design, engineering and development process. As soon as AWIS are computed on the fly from distributed data sources, the consistency and the comprehension of the delivered document is closely related to the content, the organization and the adaptation policies dedicated to users' tasks. So, a methodology is necessary to provide a global approach unifying the different components and fulfilling the user needs.

We are interested in AWIS for which users belong to a kind of community of practices: they share a common knowledge to work together [5]. We claim that the explicit knowledge of communities is the key issue to automate AWIS and thus to ensure consistency and comprehension. The specification and the management of content – **selection**, **organization** and adaptation policies – **filtering** - are separated and based on community knowledge. Thus, the AWIS design is an intensive knowledge-driven process. We have designed an AWIS as a flexible composition engine, called SCARCE - SemantiC and Adaptive Retrieval and Composition Engine based on a semantic web approach in which explicit community knowledge is formalized in ontologies. Our methodological contribution consists of: (i) a semantic organization of resources, (ii) a declarative specification of adaptation and (iii) a knowledge-driven composition engine. The main benefits are: i) a generic architecture which is reusable in different contexts, ii) this architecture is tuned to the explicit knowledge of communities and provide a method for AWIS design. Indeed, new AWIS can be generated as soon as the specification of selection, filtering or organization is modified. Of course, it is limited by the core principles underlying the composition engine. SCARCE is the core of ICCARS¹, the CANDLE² European project and KMP³.

First, some requirements for AWIS design are presented. Secondly, we develop the design principles of our architecture and methodology. Thirdly, we briefly present the adaptive composition engine. Finally, we conclude with some perspectives.

2 Some Requirements for AWIS Design

In AWIS design, content reuse leads to new requirements and constraints. We briefly analyzed some of them: interoperability and W3C standards, consistency of the delivered document, the separated management and specification of AWIS components.

Interoperability and standards: In the semantic web architecture, interoperability depends on the sharing of common architecture and standards. It takes place at three

¹ ICCARS : Integrated and Collaborative Computer Assisted Reporting System

² CANDLE : Collaborative And Network Distributed Learning Environment

³ KMP: Knowledge Management Portal, RNRT Project

different levels: i) At address level, the URL/URI referring to resources whatever their location. Indeed, information (content), programs, etc. can be resources; ii) At logical or syntactic level, where resources may be described in XML and have a DTD to explicit their logical structure. Thus, it is possible to reuse such resources with different presentations in different contexts. Nevertheless, XML resources do not have semantics understandable by machines; iii) At semantic level, where resources have metadata defining contexts, conditions and rules for retrieval and reuse. Intelligent search engines provide tools for information retrieval in using ontologies and corresponding semantic metadata associated with resources. At a semantic level, the reuse, sharing and exchange of resources are a metadata and ontology issues.

Moreover, three different views of AWIS may coexist: semantic, logical and layout [6]. Each view has a specific structure organizing it. The semantic structure of a document conveys the organization of the meaning of the content. The logical structure reflects the syntactic organization of a document. The layout view describes how the documents appear on a device. The logical and layout views are combined in an html document, whereas XML standard enable us to manage them separately. The semantic level is often implicit in traditional AWIS. New AWIS architectures, based on the semantic web, could represent explicitly the semantic level. Thus, it is possible to represent the explicit community knowledge and to share/reuse this knowledge. Indeed, such knowledge can be a resource using a standard, like OWL.

To fit the semantic web architecture, the three views can be linked to the three semantic web levels as follows: i) semantic: logic, ontology, RDFS/RDF, ii) logical: syntactic level encoded in XML, iii) layout: XSL/XSLT. New AWIS could have an architecture dealing with these three views to ensure interoperability and web standards. At runtime, the computation of the delivered document to a user will be composed of three sequential processes: semantic, logical and layout.

Separated specification and management of information system components: After having explained how to separate the presentation component in two parts: logical and layout. Now, we analyze content, adaptation and organization. Content can be reused in different contexts and different AWIS. Thus, it is necessary to dissociate organization and content and to design an information retrieval mechanism able to select the relevant content according to user needs from a user query or a content specification. Metadata schema enables designers to associate roles / functions with indexes and to enhance information retrieval. With semantic metadata, indexes have some values associated with ontologies which are formalized vocabularies. In our context, the metadata role is to automate the content reuse, whereas the LOM metadata schema [7] was not designed to automate the learning object reuse. It is difficult to automate this task, because the metadata authors, the AWIS ontologies and the end-users must share explicit and implicit knowledge. The metadata quality is crucial to retrieve the right content. There is an implicit information space corresponding to a metadata schema: all resources linked by its corresponding metadata.

Adaptation rules cannot be included or associated with content. They may not be known and under the control of the content author. As the content selection is made at semantic level through metadata and ontologies, adaptation rules have to match the user model and metadata. They also have to be consistent with the selection mechanism and the metadata schema principles. There are two issues: the granularity level

for retrieval and reuse and the use of indexes for selection and adaptation. Indeed, the retrieved resources can be reused as a whole or partially. For the latter, it is necessary to have rules and/or specific metadata entries to guide the reuse process. For the other issue, the role of indexes must be defined according to selection and adaptation.

In AWIS, organization used to be a directed graph enabling the designer to provide navigation tool and the user to access content. This graph can be explicitly defined or computed on the fly. This kind of structure can be naturally defined at semantic level by means of ontologies, thus closely related to explicit community knowledge. Content, adaptation and organization have to deal with different requirements. Moreover, they also have to face a common goal together: how to ensure consistency?

Consistency: The consistency and the comprehension of an AWIS is closely related to the content, the web site structure, presentation and the adaptation policies dedicated to users' tasks. As soon as an AWIS is computed on the fly from distributed data sources and separated views, it is more difficult to ensure the consistency of the delivered document. Thus, it is necessary to have a methodology providing a global approach unifying the different components. We could have a kind of model permitting content, organization and adaptation specifications. This model has to be able to closely link content retrieval, organization and adaptation at runtime and to ensure the global consistency.

3 Design Principles

The automation of AWIS and the content reuse can be considered as virtual documents (VD). VD are web documents for which the content, the site structure and the layout are created as needed [8]. In our approach, we consider an AWIS as an adaptive ontology-driven VD, defined as follows: an adaptive VD consists of a set of resources, their corresponding metadata, different ontologies and an adaptive composition engine which is able to **select** and to **filter** the relevant resources, to **organize** and to **assemble** them by adapting various visible aspects of the document delivered. Our adaptive composition engine consists of three composition engines: semantic, logical and layout. The selection, filtering and organization are managed at semantic level and the assembly at logical and layout level according to previous requirements. Content or reusable resources with associated metadata are called fragments. We consider two types of fragments: Atomic fragments are information units and cannot be decomposed; Abstract fragments are composed of other fragments (atomic or abstract) and one or more structures organizing them. We have a document model and an adaptation model to link content, adaptation and organization in a consistent way. These models are instantiated by an author as a generic document which specifies organization, selection and adaptation at knowledge level. The generic document is expanded and instantiated at runtime to compute on the fly a semantic graph. It is a semantic representation of the delivered document adapted to user needs.

In summary, we need a user model and a domain model and on top of that we have also a metadata schema and a document model. These models are formalized in different ontologies. They are as follows: **metadata ontology** at the information level which describes the indexing structure of resources, some index values are taken in

the domain and document ontologies; **domain ontology** representing knowledge in a specific area; **document ontology** consisting of a document model and an adaptation model; a **user ontology** which defines different stereotypes and individual features.

To begin, we present the principles underlying organization, selection and filtering.

- Organization is combined with selection and is based on a generic directed graph, having a single root, in which generic nodes have a content specification and generic arcs that are semantic relationships. The generic graph features are described in the document ontology.

- Selection is an information retrieval process on a set of resources (local or distributed) indexed with a unique metadata schema. Metadata have to be used for information retrieval and filtering which is a selection refinement. A subset of the metadata schema is used for selection specification and another one for filtering specification. The same granularity level (fragment) is used for retrieval and reuse.

- Filtering is based on adaptive navigation methods. The principle is as follows:

- First, selected resources are evaluated: the evaluation aim is to put each resource in one equivalence class according to class membership rules. These classes are declared in the document ontology according to the community of practices. Generally, a maximum of five equivalence classes are defined. It might be difficult for a user to deal with too many equivalence classes [1, 9].

- Second, one adaptive technique is chosen for the current user, its preferences and the stereotypes associated with adaptive navigation techniques. Thus, it is applied to delete, annotate, hide, etc. some equivalence classes of resources.

The Document Ontology: It is composed of a document model, an adaptation model, subcategories and instances of fragments, semantic relationships and generic documents. The two models belong to the core of SCARCE whereas the others depend on the community of practices. They are parts of the common knowledge shared by community members. The document model is based on the concept of *Abstract Fragment* (fig. 1). In ICCARS and CANDLE, we chose to associate a particular information space with an abstract fragment to enhance AWIS consistency. Nevertheless, this constraint can be released and all fragments indexed by the metadata schema could be used. It depends on the communities of practices.

An abstract fragment can be organized according to one or more *Structures*. A structure is a collection of *Generic Nodes* among which one is its root. A generic node is an abstract object, which only exists inside one structure. It is linked to others through a *Semantic Relation*. Semantic relations are specific to the community of practices (such as RST [10]). The set composed of generic nodes and the corresponding relationships is one organization of the AWIS. An abstract fragment is a directed graph in which the nodes are generic nodes and the vertices are semantic relations between generic nodes. At runtime, a generic node can be view as an information retrieval process which uses a description given by the author according to metadata, to send a query to the intelligent information broker. It is able to use the user model to filter the small set of resources. Generic documents can be organized in

a hierarchy in which classes are patterns of generic documents. In ICCARS and CANDLE, organizations are narrative structures which represent the author competences and know-how [11]. In KMP, organization is close to a task model defining practice scenarios.

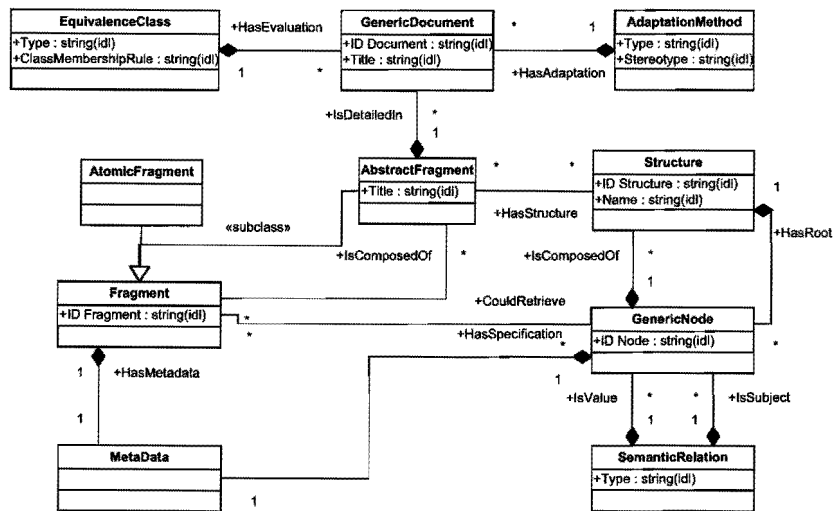


Fig. 1. Document Model

The adaptation model is composed of the equivalence classes, their class membership rules, the set of adaptation methods and their stereotypes. A generic document is also composed of adaptation rules which are instances of the different elements of the adaptation model. A class membership rule is a comparison between user’s characteristics (user ontology), and fragment’s indexes (metadata ontology). An adaptation method is permitted when a user fits the stereotype associated to that method.

The Metadata Ontology: The metadata schema (Table 1) provides metadata information about fragments. The semantic composition engine uses the schema for information retrieval. This schema is composed of two kinds of characteristics, typical entries for web resources such as the author, the language and the date of creation, and more specific entries that depends on the community of practices. The typical entries are usually found in numerous metadata schemas, but they are not compulsory, it depends of the application. A gray background shows the specific entries.

Table 1. The metadata schema of ICCARS

MD.1	General	General information about the resource	
MD.2	Life cycle	Entries for versioning purposes	
MD.3	Meta Metadata	Information about metadata	
MD.4	Technical	Technical information about the resource	
MD.4.1	Location	Where the resource can be found?	
MD.4.2	Format	Format of the resource	
MD.4.2.1	Type	Type of the resource (ppt, doc, html, ...)	
MD.4.2.2	Size	Size of the resource in Kbytes	

MD.5	Classification	Data about the content and reporting features	
MD.5.1	Domain	Description related to the domain	
MD.5.1.1	Concept	Concept name	Domain ontology
MD.5.1.2	Level	Level of knowledge required	
MD.5.2	Reporting	Reporting classification	
MD.5.2.1	Resource Type	Type of resource (Interview, report ...)	Document ontology
MD.5.2.2	Edition	Edition concerned	Document ontology
MD.5.2.3	City	City concerned	
MD.6	Rights	Use conditions of the resource	

That schema is composed of six sections, typical sections are usually proposed in all metadata schemas ("General", "Technical", ...). Some sections are more specific to our architecture, it is the case of "LifeCycle" which is necessary for the management of versioning purposes [12]. "Classification" has to describe the content of fragment according to the domain model and to the community of practices.

The User Ontology: It describes the user characteristics for adaptation purposes. It is used by the adaptation model to define the class membership rules and the stereotypes of adaptive methods. The user ontology needs the domain ontology to describe the knowledge of the user – as an overlay model. Generally, the user model consists of the following five parts: personal, preferences, knowledge, history and session. It is possible to add some specific sections according to the community of practices. "Personal" is composed of typical data on the user such as his name. "Preferences" is useful for personalization purposes; it is a section that can be modified easily by the user. "Knowledge" describes the overlay model, and is used by the system for evaluating the relevance of fragments. "History" is used for proposing facilities to the user inside the environment such as bookmarks. "Session" is used only by the system as a storage area for the current session. "History" and "Session" are specific to our architecture because they are used by our composition engines.

Table 2. The user model of ICCARS

UM.1	Personal data	Personal data concerning the user	
UM.1.1	Identity	His identity	
UM.1.2	Login	Unique identifying data	
UM.1.3	Classification	Classification data	
UM.1.3.1	Location	Where does he live?	
UM.1.3.2	Professional Activity	What kind of job? (economist, fisherman, student, etc.)	
UM.1.3.3	Role	The role in the application (author, reader ...)	
UM.2	Preferences data	Data about the preference of the user	
UM.2.1	Interest	Topics of interests	
UM.2.1.2	Topic	A list of topics	Domain ontology
UM.2.2	Adaptation	Adaptation preferences	
UM.2.2.1	Element	An element which can be adapted (link ...)	
UM.2.2.2	Rule	A method of adaptation (annotation ...)	Document ontology
UM.3	Knowledge	Data about the knowledge of the user	
UM.3.1	Domain	Knowledge about the domain	
UM.3.1.1	Element	A domain concept	Domain ontology
UM.3.1.2	Level	A level of knowledge	
UM.4	History data	Data about access to generic documents	
UM.5	Session data	Data concerning the current session	

The Domain Ontology: It is a typical domain model which is composed of a semantic network for the domain of the application. It is the link between the user knowledge, the fragment description and the generic node content specification.

4 Adaptive Composition Engine

The four mechanisms of VD are implemented as follows: selection, filtering and organization are achieved in the semantic composition; assembly is divided into logical and layout compositions. The generation of the document is detailed in [13].

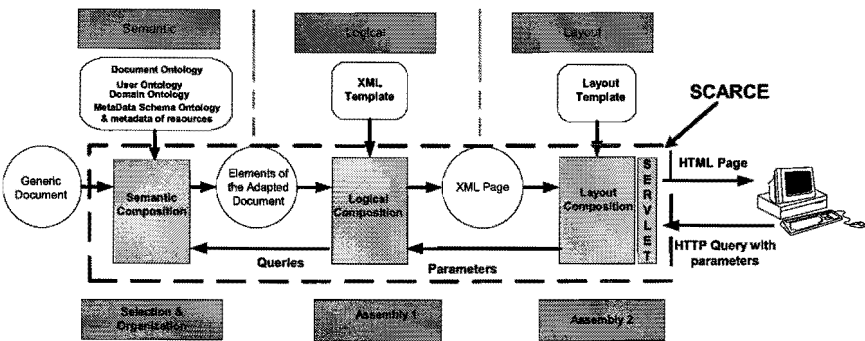


Fig. 2. The Composition Engine Architecture

The semantic composition engine computes on the fly a **semantic graph** from a **generic document**. This semantic graph represents a narrative structure adapted to the current user and linked to its corresponding content. The aim of the logical and layout composition engines is to convert the semantic graph into a real document. This process depends on the nature of the real document. The navigation and the resource access are based on the semantic graph. The logical composition engine requests the semantic composition engine to build and to browse the semantic graph according to user interactions and computes for each node an XML web page with its content and navigation tools. The layout composition engine generates an HTML page from the XML web page by applying the layout rules.

Now, we present briefly the request-response cycle of the composition engine. The real document is computed on the fly stage by stage. A stage consists of the computation of the next web page according to user interaction. An XML template describes a generic web page which is composed of several dynamic components (navigation guides, etc). For instance, we assume that we have local and global navigation guides and content inside a template. In our example, we focus on the local navigation guide which is paradigmatic (fig. 3.). In this figure, the local navigation guide consists of three types of links: previous, current and next. It applies the annotation technique: one or several stars are associated with a link. To be able to choose the right number of stars, it is necessary to evaluate the corresponding resources and to classify them in an equivalence class. All the resources are associated with a node of the semantic graph and then with the corresponding node of the generic directed graph. The "next" links enable the user to select the direct neighbors of the current web page. They consist of all the next nodes in the semantic graph. Therefore, we have to compute in advance the direct neighbors in the semantic graph before displaying this component.

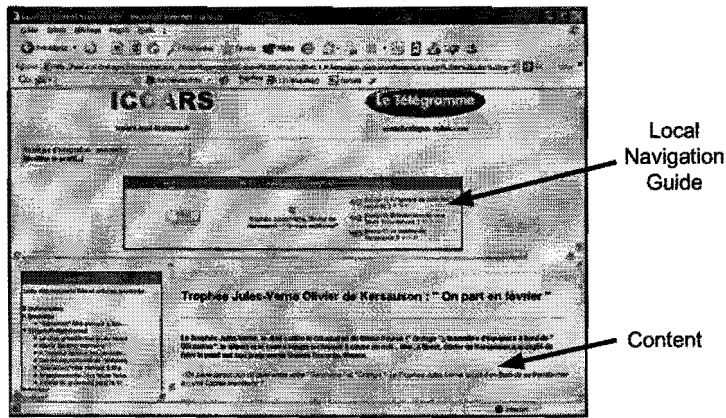


Fig. 3. Web Page layout

From a user interaction, the servlet engine receives an HTTP query with parameters: the component of the query and the names of the XML template, the generic document, the next node and the content of the next web page. The layout composition engine sends these parameters to the logical composition engine. It accesses the XML template with several XML components. For each XML component, a java component requests the semantic composition engine to instantiate its XML component and generates an XML structure. For instance, the local navigation guide component asks the semantic composition engine to compute the previous, current and next web pages of the new one and their equivalence classes. Then, it is able to get an XML structure having these types of links with their parameters and equivalence classes. From these computations, the logical composition engine is able to instantiate the entire XML template and to obtain an XML web page. Then, the layout composition engine generates the web page from the corresponding layout template.

5 Conclusion

In this paper, we have presented a methodology and an architecture for the content reuse in the design of AWIS. It ensures interoperability in separating the specification and the management of resources, organization, adaptation and layout. Content, organization and adaptation are specified by an author according to its explicit community knowledge. Our methodological approach relies on: (i) a semantic organization of resources by means of the document model and the community know-how, (ii) a declarative specification of the generic document at knowledge level, (iii) a knowledge-driven engines for the composition of AWIS. We propose a generic architecture for AWIS that have to be instantiated for specific communities of practices. Our approach is limited by the core principles of the architecture. To enhance the consistency of the delivered document, we chose to associate an information space to an abstract fragment. The main idea is to ensure that we have the right content at the

right place for the right user. It will be interesting to release this constraint according to the user goals and the services we want to propose. In the case of a typical information retrieval system, it is necessary to deliver the maximum of relevant fragment, but in the case of a pedagogical application, it is necessary to ensure the coherence between all fragments retrieved. We don't have a solution at the moment; it is an interesting problem that has to be studied in the future. SCARCE engines are currently implemented in Java and linked to the servlet engine. In ICCARS and CANDLE, the ontology-based inference engine used is Ontobroker [14]. In KMP, we are developing a new interface in the semantic composition engine for Corese [15].

References

1. Brusilovsky, P., *Methods and techniques of adaptive hypermedia*. User Modeling and User-Adapted Interaction, 1996. 6(2-3): p. 87-129.
2. Aroyo, L., P. De Bra, and G.-J. Houben. *Embedding Information Retrieval in Adaptive Hypermedia: IR meets AHA!* in *AH 2003: Workshop on Adaptive Hypermedia And Adaptive Web-Based Systems*. 2003: Technische Universiteit Eindhoven.
3. Brusilovsky, P. and H. Su. *Adaptive Visualization Component of a distributed Web-based Adaptive Educational System*. in *ITS 2002*. 2002. Biarritz: Springer Verlag.
4. Conlan, O., et al. *Applying Adaptive Hypermedia Techniques to Semantic Web Service Composition*. in *Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems in The Twelfth International World Wide Web Conference*. 2003. Budapest, Hungary.
5. Wenger, E., *Communities of Practice - Learning, Meaning and Identity*, in *Learning in Doing: Social, Cognitive and Computational Perspectives*. 1998, Cambridge University.
6. Christophides, V., *Electronic Document Management Systems*. 1998, UCH/FORTH: <http://www.ics.forth.gr/~christop/>.
7. LTSC, *Learning Object Metadata Working draft document 3*. 1999, Learning Technology Standard Committee: <http://ltsc.ieee.org/doc/wg12/LOM-WD3.htm>.
8. Milosavljevic, M., F. Vitali, and C. Watters. *Introduction of the virtual document workshop*. in *Workshop on Virtual Document, Hypertext Functionality and the Web*. 1999. Toronto.
9. De Bra, P. and L. Calvi, *AHA! An open Adaptive Hypermedia Architecture*. The New Review of Hypermedia and Multimedia, 1998. 4: p. 115-139.
10. Mann, W.C. and S.A. Thompson, *Rhetorical Structure Theory: Toward a functional theory of text organization*. Text, 1988. 8(3): p. 243-281.
11. Iksal, S., et al., *Semantic composition of special reports on the Web: A cognitive approach*, in *H2PTM'01*, P. Balpe, et al., Editors. 2001, Hermès. p. 363-378.
12. Iksal, S. and S. Garlatti, *Revisiting and Versioning in Virtual Special Reports*, in *Hypermedia: Openness, Structural Awareness and Adaptivity, Third Workshop on Adaptive Hypertext and Hypermedia*, S. Reich, M.M. Tzagarakis, and P.D. Bra, Editors. 2001, Springer Verlag: Århus, Denmark. p. 264-279.
13. Garlatti, S., S. Iksal, and P. Tanguy, *SCARCE: an Adaptive Hypermedia Environment Based on Virtual Documents and the Semantic Web*, in *Adaptable and Adaptive Hypermedia Systems*, S.Y. Chen and G.D. Magoulas, Editors. 2004, Idea Group. p. To be appear.
14. Fensel, D., et al. *On2broker in a Nutshell*. in *the 8th World Wide Web Conference*. 1999.
15. Corby, O. and C. Faron-Zucker. *Corese: A Corporate Semantic Web Engine*. in *Workshop on Real World RDF and Semantic Web Applications 11th International World Wide Web Conference*. 2002. Hawaii.

Effectively Deploying an Adaptive Web Based eLearning System by Extending UML

Emma Quinn, Michael O'hEigartaigh, Pramod Pathak

National College of Ireland, Mayor St. IFSC, Dublin 1, Ireland
{equinn, moheigartaigh, ppathak}@ncirl.ie

Abstract.

eLearning is increasingly being integrated into organisations and universities as a new means of teaching and supporting learners. However, eLearning is still in its infancy; migrating from a classroom-training environment to an e-learning environment is a major undertaking for any organization. We propose the use of an extended UML methodology for the development and deployment of eLearning in an organization to support migration between environments. It is proposed to use the methodology in a controlled environment involving Java students who use a drop-in centre in a third level institution.

Introduction

eLearning is increasingly being integrated into organisations and universities as a new means of teaching and supporting learners. In April 2004, Sheffield University upgraded its virtual environment to meet the growing demand of its 12,000 students. [1]. In most cases eLearning is used to support instruction led classes in a blended learning environment; however, eLearning is also sometimes used as the sole method of instruction.

Moving from a classroom-training environment to an eLearning environment is a vast undertaking within an organization both for employees and managers alike. The transition involves migrating training functions or training components from a static, physical environment to a dynamic, virtual environment. This, in itself, causes a substantial change in the training function and in the technical infrastructure of the organization. Therefore, it is necessary that this transition is planned, controlled and coordinated, prior to, during and following its migration. The scale of complexity of the problem has many similarities with software development but it also has to encompass a cognitive dimension based on learning.

The research paper has one primary goal:

1. Can UML be extended as a paradigm for eLearning in organisations?

We propose the use of an extended UML methodology for the development and deployment of eLearning in an organisation. UML is utilized by many organisations to either plan for introducing a new system or to plan for upgrading an existing system.

There are a number of different issues to consider when migrating from classroom learning to online learning. The virtual learning environment forms the basis of the eLearning system. It encompasses and affects all elements that are unique to eLearning, such as, the use of technology, (non) human interaction and so on.

Technology is the medium through which eLearning content is transmitted. Today, the use of technology is becoming the norm in everyday working life. According to EuroStat, in 2002, more than 50% of over 15 year olds in Europe are Internet users [2]. The following table lists statistics from 2002 based on small, medium and large Canadian organizations.

Table 1. Survey of Electronic Commerce and Technology for 2002

Type	Small Company	Medium Company	Large Company
Use of PCs	84%	97%	100%
Use of Internet	73%	92%	99%
Company Website	27%	62%	77%
High Speed Access	56%	71%	84%

Blended Learning Model

The Blended Learning Model enables a trainer/lecturer to chart which learning objects are to be used in the classroom environment and which objects are to be used in the virtual environment.

The same learning object is applied to the virtual environment as well as to the classroom environment. The model enables the trainer/lecturer to effectively plan a blended learning module

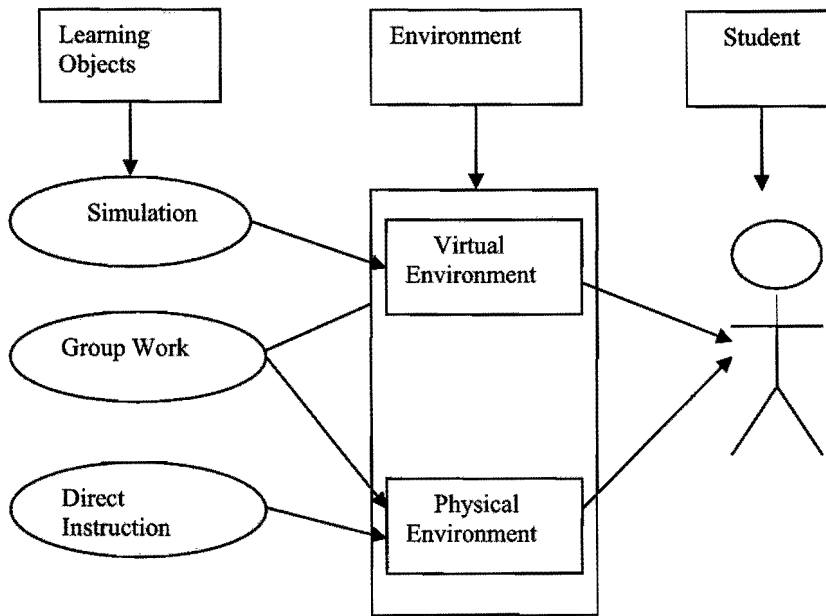


Fig. 1. Blended Learning Model

Learning Object Model

Peter E. Doolittle categorizes a series of instructional strategies to consider when designing an eLearning course [3]. According to Dick, Carey and Carey, an instructional strategy encompasses the methods, material, content and processes specified in the learning activities [4]. Doolittle's categorization of instructional strategies includes, for example, synectics, simulations and questioning & instruction. This section attempts to model the instructional strategies defined by Dick Carey and Carey and as categorized by Doolittle in terms of instructional objects in UML.

The model represents instructional strategies as instructional objects. The instructional objects incorporate elements such as learning object, assessment object and delivery method. Class diagrams in UML model objects in software engineering, including an object's state, function and relationship. Class diagrams can be extended to model an instructional strategy as an instructional object. Similar to object-oriented software, the instructional objects have states, behaviours and relationships with other objects.

However, a learning object should contain a metric for evaluating how effectively it matches the cognitive state of the student. In a simple model, Bloom's taxonomy could form the basis of the comparison. Thus a set of weights could be attached to an eLearning object covering knowledge, understanding, translation, analysis, synthesis and evaluation. If an eLearning object is prepared in a number of representations

relative to this weighting, the formulation that best meets the profile of a selected student can then be selected. In our model, the coefficient of correlation was used as a metric. This has the advantage of being well understood and is invariant to linear changes in the scale of measurement of the variables.

Correlation Model
Objective: Modelling for a Java lesson aimed at first year college students

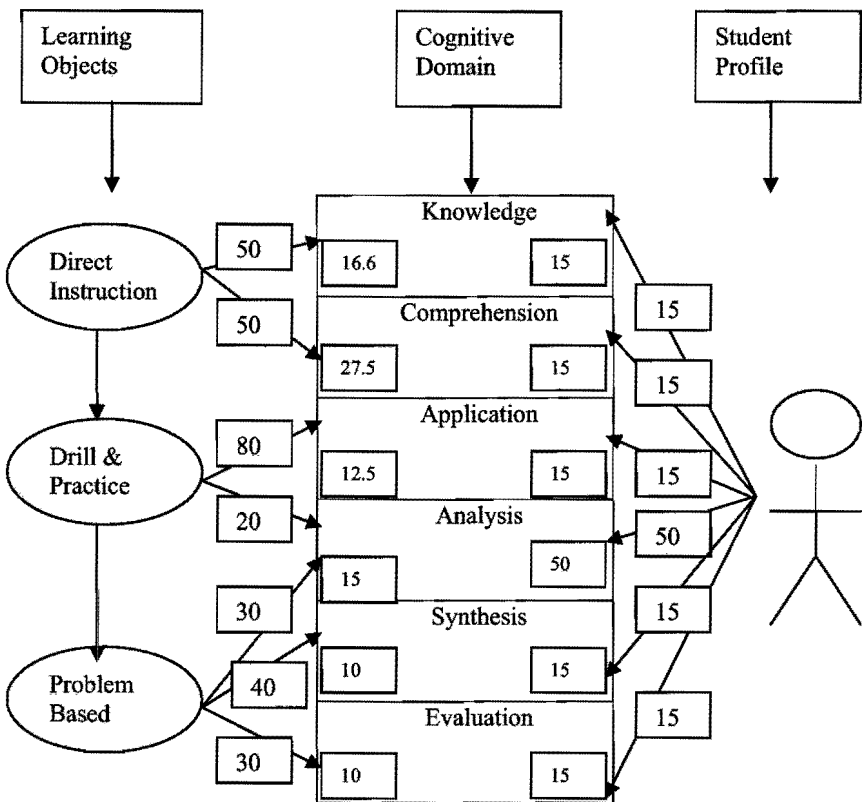


Fig. 2. Correlation Model

Table 2. Correlation Table

Cognitive Domain	Total Module Spent At each Cognitive Domain Level	Average Module	Total Student Profile	Correlation
Knowledge	50	16.6%	15%	-1.6%
Comprehension	50	16.6%	15%	-16%
Application	80	26.6%	45%	18.4%
Analysis	30	10%	10%	0%
Synthesis	40	13.3%	10%	-3.3%
Evaluation	30	10%	5%	-5%

Many organisations and academia stress the existence of challenges that must be overcome in order to successfully implement eLearning. "The difficulties facing the educational practitioner are twofold: (i) becoming familiar with the complexities of the new pedagogical approaches allied to a new and very fast moving technological environment and (ii) being able to identify the relevance of these new approaches in his/her own daily practice" [5] . "It is highly unlikely, according to industry statistics that the implementation will occur without obstacles" [6].

Conclusion

Implementing and extending the UML model should be a primary focus for further research. This paper presented a brief overview of potential implementations. Further research must be carried out if extended UML models are to be implemented efficiently. The models in this paper apply Bloom's Taxonomy; a possible research direction would be to apply a variety of learning theories to the models.

The model will be tested on students who attend a drop-in centre for Java programming in an Irish third level institution. This will require that the learning material be prepared in a suitable form. It is hoped that the overhead involved in this will be offset by the enhanced learning experience of the students. We shall report the findings of the study in a future publication.

References

1. Mortelman J.: Sheffield Uni Joins eLearning Converts (2004) Available online at: <http://www.vnunet.com/news/1154020>

2. Measuring the Information Society-Eurostat activities and Action Plan for Information Society Statistics, presentation at the Labour Market Statistics Conference (2002)
3. Doolittle P.E.: Instruction Design for Web-Based Instruction, Second Annual Irish Educational Technology Users Conference (2001)
4. Dick, W., Carey, L., Carey J.O.: The Systematic Design of Instruction. 5th Edition. Addison-Wesley Educational Publishers (2001)
5. Nunes, J.M; McPherson M.A.: No Lectures On-Campus: Can eLearning Provide a Better Learning Experience? Proceedings of the 2nd IEEE International Conference on Advanced Learning Technologies (2002)
6. Intelera White Paper Software Implementation Challenges and Solutions

Navigational Properties and User Attributes for Modelling Adaptive Web Applications

Gonzalo Rojas¹, Vicente Pelechano², and Joan Fons²

¹Departamento de Ingeniería Informática y Ciencias de la Computación
Universidad de Concepción
Edmundo Larenas 215
40392 Concepción, Chile
grojas@inf.udec.cl

²Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia
Camino de Vera s/n
46022 Valencia, Spain
{pele, jjfons}@dsic.upv.es

Abstract. This work presents an approach for specifying the navigational characteristics of Adaptive Hypermedia Systems, providing the conceptual primitives of a pre-existing Navigational Model with adaptive navigational properties. These properties allow managing the accessibility and order of the modelled navigational paths according to the characteristics of the user. Values of these properties define the navigational alternatives that modelled application offers to the different kinds of users. The effects of the introduced properties on the final Web application are shown by means of an example of use. This work is part of a global proposal for Adaptive Web Systems modelling, in the context of the OOWS approach.

1 Introduction

Traditional Web applications provide users with a limited set of interaction possibilities, with few differences about navigational alternatives, contents of data items and presentational features. In this context, Adaptive Hypermedia arises as an approach that allows achieving usage experiences that are more suitable to the individual information needs. Goals, preferences and knowledge of individual users are described in a User Model and are used for adaptation to their needs [1].

It is necessary to face the challenge of Web adaptive applications modelling from a high level of abstraction, which allows the modeller describing the different users of the application and the sets of interactions that will be provided to each of them.

Some Web development approaches as WebML [2], OO-HDM [3], OO-H [4] and UWE [5] present global proposals for tackling the navigational modelling of adaptive Web applications. However, most proposals are limited to the definition of constraints over the navigational model of a non-adaptive application.

In this work, we introduce some properties that allow implementing adaptive navigation techniques, such as *link-hiding* and *link-ordering* [1]. Adaptive properties are defined over pre-existent conceptual primitives of the OOWS Navigational Model [6]. This proposal is framed within a global proposal of Adaptive Web applications modelling, which is based on correspondences between semantic descriptions of users and applications.

In Section 2, we present the adaptive properties that are defined over conceptual primitives of OOWS Navigational Model, including an example that shows the effects of these properties on the final application. Finally, section 3 shows some conclusions and related works that complement this proposal.

2 Supporting Adaptive Navigation in OOWS

OOWS (Object-Oriented Web Solution) modelling method is the extension of the OO-Method approach [7] that allows developing web applications. The OOWS Navigational Model introduces the required expressiveness to capture the navigational requirements of web applications.

A navigational map is defined for each kind of user. This map is represented by a directed graph in which the nodes are *Navigational Contexts* and the arcs are *Navigational Links* that defines the valid navigation paths.

Navigational contexts represent user interaction points that provide a set of cohesive data and operations. Three types of navigational contexts are defined:

- *Exploration context*, which is reachable from any node.
- *Sequence context*, which can only be accessed via a predefined navigational path.
- *Mixed context*, which can behave as an exploration or as a sequence context, depending on the user that currently accesses to it.

A Navigational Context is composed by *Navigational Classes*, which define views of the object model classes. Each navigational context has one mandatory navigational class from which the information retrieval starts, called *manager class*, and other optional navigational classes that provide complementary information, called *complementary classes*.

All navigational classes must be related by a unidirectional binary relationship (*Contextual Relationship*). It is defined over an association, aggregation, composition or specialization relationship included in the object model. Contextual relationships of three types are defined:

- *Complementary Information relationship* (graphically represented using dashed arrows), which represents a basic information recovery by crossing a structural relationship between classes.
- *Navigational relationship* (solid arrows), which represents an information recovery and navigation to a target node.
- *Mixed relationship* (dash-point arrows), which can be a Navigational or a Complementary Information relationship, depending on the current user.

Figure 1 shows the *Books* Navigational Context, which describes a page of a specific book in a web-based bookstore. *Book* manager class shows basic purchase data of the presented book, while information from complementary classes is retrieved through the shown navigational relationships.

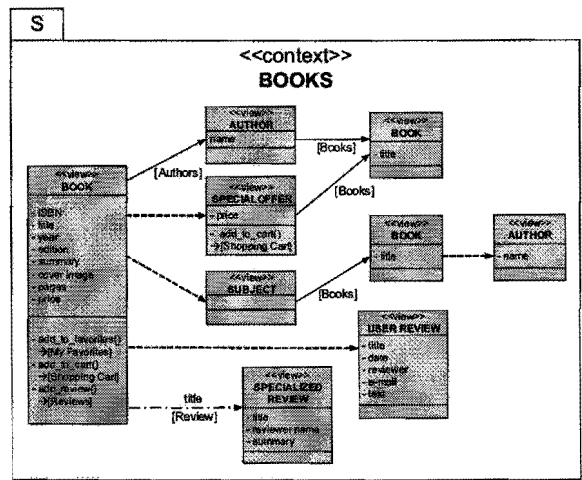


Fig. 2. *Books* OOWS Navigational Context for a Web-based bookstore.

2.1 Accessibility property for Link-Hiding

This approach considers the link-hiding technique at conceptual level, by means of the definition of the Accessibility property over navigational contexts. Any context will be “*accessible*” or “*not accessible*”, depending on the user’s characteristics. This property allows managing the adaptation of the access to contents that are distributed among different navigational contexts. In the example of the web bookstore, an exploration context that allows users registering as clients may be “*accessible*” only for users that are not already registered.

The Accessibility property is also defined over contextual relationships. Any contextual relationship can be “*accessible*” or “*not accessible*” for certain users. In this case, the accessibility property allows adapting the access to contents that are included within a given context. For example, the relationship that allows accessing a special offer associated with a book will be “*accessible*” only for clients with a certain amount on previous purchases. The accessibility property of contextual relationships implements a link-hiding technique, but also manages the access to contents with no involved navigation, in the case of complementary information and mixed relationships.

2.2 Importance Weight for Link-Ordering

This proposal establishes a contextual relationship ordering based on importance weights, with the purpose of privileging the data and navigational paths that are most relevant for a given user. Contextual relationships whose source is the manager class are assigned with a relative weight, establishing an order of importance among these relationships for a specific user type. According to user characteristics, displaying of information retrieved from one or another relationship is prioritized.

As an example, let us consider a client looking for a book in the mentioned Web bookstore to give his mother as a birthday gift. This client and his mother are not regular readers, and probably he is looking for something easy-to-read. He is not concerned about the price of the chosen book.

Information about the characteristics and needs of this client is obtained through the use of some technique for user information retrieval [8] and is described by means of values of user attributes. These values define a certain numerical weight for each contextual relationship of a given context. In this way, considering subsets of context relationships with the same source navigational class, along with the set of relevant attribute values describing a specific user, a score for those relationships is obtained. This score allows ordering the contextual relationships of that context for this user.

By applying this technique to *Books* navigational context of Figure 1, one possible order of its context relationships, for the described specific client and in decreasing order, is the following:

1. *Book-Author*: probably most users privilege the relevance of data about the book's author(s); our example's client should be interested, at least, on his/her name.
2. *Book-User Review*: irregular reader characteristic may be a user attribute value that raises the weight of this relationship. An occasional reader can better estimate the other readers' opinion, probably comparing it to his/her mother's preferences.
3. *Book-Subject*: a "purchase reason" attribute (in this case, "mother's birthday"), along the mentioned non-habitual reading characteristic, may increase this relationship's weight, allowing user to select among other similar alternatives.
4. *Specialized Review*: because of similar reasons, this relationship is not as important for this user as the *Book-User Review*.
5. *Book-Special Offer*: the "purchase reason" and the little importance of book price for the user may justify a low weight for this relationship.

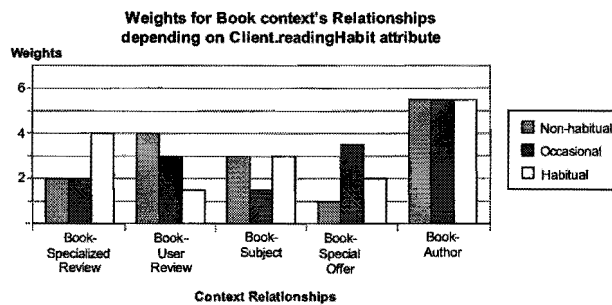


Fig. 5. Weights for Contextual Relationships depending on a user attribute.

Figure 2 shows the weights that the modeller assigned to the relationships of *Books* context. Depending on the *Client.readingHabit* attribute value, different weights for those context relationships are assigned.

Two possible implementations of the *Book* Navigational Context are shown in Figure 3. We can distinguish the main data of the presented book, corresponding to attributes and operations of the manager class *Book* (1, in both sides); the data retrieved through *Book-Special Offer* relationship (2, left side) that show a special offer related to the book; the book review by a specialist, through *Book-Specialized Review* (3); the other clients' book reviews, through *Book-User Review* (4); other books by the same author, through *Book-Author* (5); and other related books through *Book-Subject* (6).

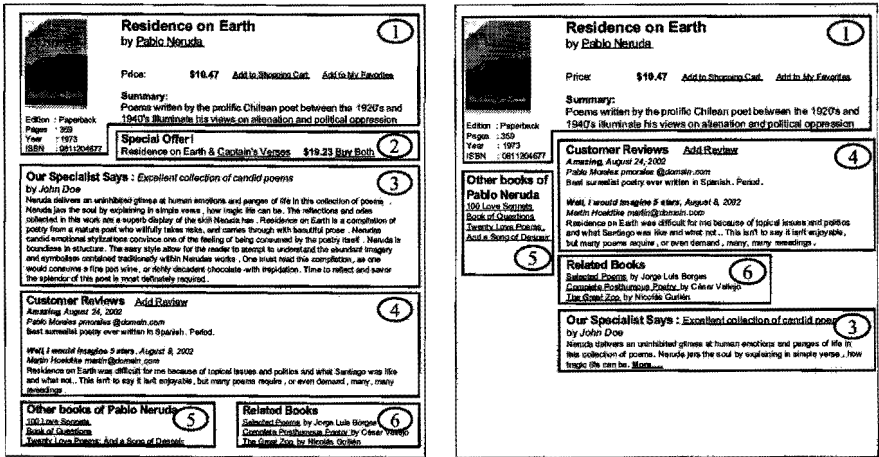


Fig. 6. Possible implementations of *Book* Navigational Context, without (left) and with (right) adaptive navigation properties (information extracted from Amazon.com)

The left side of Figure 3 shows an implementation of *Book* context, where adaptive properties have not been considered. The resulting Web page is shown with no differences for all types of user. The right side shows the final Web page that is presented to the client of the mother's birthday present, considering the described adaptive properties. It shows the same information items and operations of the other page, but in this case they are ordered according to the scores that were obtained for each contextual relationship. Let us suppose that special offers are only shown to clients with a certain amount on previous purchases on this site and that this is not the case of this client. *Book-Special Offer* navigational relationship is "not accessible" for him, so the information corresponding to (2) is not included in the final Web page.

3 Conclusions and Future Work

This work presents an approach for achieving high-level descriptions of navigational features for Adaptive Hypermedia Systems. Our proposal introduces adaptive properties over navigational structures, augmenting the expressiveness of a particular navigational model, avoiding the definition of new conceptual primitives. Well-known adaptive navigation techniques, such as link-hiding and link-ordering, are considered into the application modelling, with little effort demanded to the modeller.

Current works face the adaptive navigation by incorporating constraints over navigational models of non-adaptive Web applications. Our proposal considers the adaptive features as an inherent aspect of any Web application, incorporating adaptive properties as an essential feature of the navigational modelling process.

We are working on the following related topics: (1) adaptation of accessibility of navigational class operations; (2) consideration of the semantic distance among contextual relationships; (3) definition of adaptive constraints over contextual relationships; and (4) definition of adaptive properties for OOWS presentational model.

References

1. Brusilovsky, P.: Methods and Techniques of Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, N°6 (1996) 87-129.
2. Ceri, S., Fraternali, P., Paraboschi, S.: Data-Driven One-to-One Web Site Generation for Data-Intensive Applications. *Proceedings of the 25th VLDB Conference*, Edinburgh, Scotland (1999) 615-626.
3. Rossi, G., Schwabe, D., Guimarães, R.: Designing Personalized Web Applications. *Proceedings of the World Wide Web Conference (WWW'10)*, Hong Kong, China (2000) 275-284.
4. Garrigós, I., Gómez, J., Cachero, C.: Modelling Dynamic Personalization in Web Applications. *Third International Conference on Web Engineering (ICWE'03)*. LNCS, Vol. 2722. Oviedo, Spain (2003) 472-475.
5. Koch N., Wirsing, M.: Software Engineering for Adaptive Hypermedia Applications? *Third Workshop on Adaptive Hypertext and Hypermedia*. 8th International Conference on User Modeling (2001).
6. Fons, J., Pelechano, V., Albert, M., Pastor, O.: Development of Web Applications from Web Enhanced Conceptual Schemas. *Proc. of the 22nd International Conference on Conceptual Modelling, ER'03*. LNCS, Vol. 2813. Chicago (2003) 232-245.
7. Pastor, O., Gómez, J., Insfrán, E., Pelechano, V.: The OO-Method Approach for Information Systems Modelling: From Object-Oriented Conceptual Modelling to Automated Programming. *Information Systems*, Vol. 26, N. 7 (2001) 507-534.
8. M. Eirinaki, M. Vazirgiannis, M.: Web Mining for Web Personalization. *ACM Transactions on Internet Technology*, Vol.3-1 (2003) 1-27.

Plasticity in Mobile Devices: a Dichotomic and Semantic View

Montserrat Sendín¹, Jesús Lorés¹

¹ Computer Science and Industrial Engineering Department
University of Lleida, 69, Jaume II St., 25001- Lleida, SPAIN
Tel: +34 973 70 2 737 Fax: +34 973 702 702
{msendin,jesus}@eup.udl.es

Abstract. Everybody knows that mobile computing provides us with tremendous versatility. However, the more this versatility increases the more challenging the goals and demands of plasticity becomes. On the one hand, the complexity in the design of User Interfaces (henceforth UIs) increases, and on the other hand the need for dynamic adaptation becomes more noticeable. In this paper we analyze the problem of plasticity of UIs and we distinguish and define two levels of actuation in the plasticity process associated to two new subconcepts we name explicit plasticity and implicit plasticity respectively with two different goals. This paper also revises a client/server-based architectural framework proposal -already presented in [14]- focussing on the implicit plasticity problem. On the one hand, it proposes an adaptive architecture to accommodate this continuous and diverse variability in mobile devices (client side) based on the combination of reflection techniques and Aspect Oriented Programming (AOP) paradigm [11]. In the server side it is also revised the plastic UIs development framework, and presented a semantic approach. The aim is to provide the necessary infrastructure to guarantee the plasticity of UIs contributing at the same time to reusability and flexibility, and adjusting in short to the canons of quality of a software product.

1 Introduction

Today technology allows users to keep on moving with computing power and network resources at hand. Computers are shrinking while the bandwidth of wireless communications keeps increasing. These changes have increasingly enabled access to information “anytime and anywhere”, making computing possible in multiple and varied *contexts of use* in which multiple variable parameters come together. This provides us tremendous versatility. Nevertheless, *runtime constraints* related not only to resources constraints (bandwidth, server availability, physical resources restrictions, etc.), but also related to the user (user’s changing needs, tasks to be developed, profile and current situation), and even related to the environment (location, day of the week, hour, weather conditions, etc.), are volatile and require sophisticated adaptive capabilities that today are still challenging. Today systems should be prepared to face up this continuous and diverse variability, offering the capacity of systematically pro-

ducing as many UIs as contextual circumstances have been identified. This fact raises a first challenge. But apart from that, the adaptive capacities should evolve in runtime, as the real time constraints vary, offering also continuous adaptation to each concrete UI in use, with the aim of solving the contextual changes in a dynamic and automatic way. We identify this problem as a second challenge different from the previous one.

Until now these two issues have been designated a single term: *plasticity of UIs*¹, introduced by Thevenin and Coutaz in [19] in 1999 along a framework and research agenda development work. This work was put in practice in these 2 works: ARTStudio [18] and Probe [4], and revised in 2002 in [5], within the Cameleon project [6].

According to the distinction we have glimpsed at and explained above, we present an extension to the Thevenin and Coutaz work consisting of distinguishing and delimiting these two levels of operation in the plasticity process which give rise to two subconcepts of plasticity that we have named *explicit plasticity* and *implicit plasticity*. They have been defined in detail and set bounds to the proposed goal in each case in [14].

As it appears in [14], we define the *explicit plasticity* as the capacity of automatically generating a specific UI valid to a concrete *context of use*², starting from an abstract specification -*generic UI*³. We define *implicit plasticity* as the capacity of incremental adaptation that the specific UI obtained should adopt in real time as the user goes across new contexts of use. The *explicit plasticity* tackles important changes in the UI that are caused either by unforeseen situations –a change in the user or the device- or by the request of new contents –changes in the user location or in the task to perform-, that in any case involve a reconfiguration of the UI (this case also includes the initial configuration of the UI), to solve in the server side upon explicit request. This is the reason why we use the term *explicit*. The *implicit plasticity* tackles specific modifications in the UI, originated by predictable contextual changes (changes in the brightness level, server connectivity level, bandwidth, small advances made by the user, etc.), which can be solved with an automatic local readjustment on the client side, without any express request. This is why we chose the term *implicit*. Moreover, both concepts are located in different runtime moments, and also are solved in the opposed sides of a client/server-based architecture. The explicit plasticity is solved in the server side, and the implicit plasticity in the client side. This fact means that each one requires different architectural frameworks, strategies, modeling and implementation techniques.

In the case of the *explicit plasticity* it is required a systematic development support capable of generating the UI suitable to the posed cases, avoiding having to manually

¹ Capacity of a same generic UI to support variations in multiple contextual circumstances preserving usability and at the same time minimizing the development and maintenance costs [6]. This term comprises diverse properties: user adaptivity, context-awareness and appliance-independence, preserving at the same time usability.

² Set of environmental parameters that describe a particular context in which a set of conditions of real time come together, related not only to the restrictions in hardware resources, but also related to the user, and even to the environment.

³ Unique and flexible abstract specification of the layout and composition of the resulting interface, as well as a description of how the UI evolves in time. It is related to the concept of canonical expression of the UI.

develop as many UIs as possible contexts of use. We call this support *explicit plasticity engine*. In the case of the *implicit plasticity* an architecture with evolving capacity that allows easily to adapt the UI to the contextual changes is required. We are referring to an engine, as flexible as possible, capable of detecting the environment and of appropriately reacting, applying the necessary adjustments to the UI. We call this architecture *implicit plasticity engine*. This paper focuses more on the implicit plasticity concern and proposes an architecture based in the combination of reflection techniques and AOP paradigm. Moreover, the paper also presents a general description for the explicit plasticity engine we propose in the server side.

This paper is composed of three sections. In section two it is featured the implicit plasticity term giving an overview. Then, in section three we present our client/server-based architectural framework proposal. On the server side we present our semantic approach which consists of adding a conceptual layer to the development framework presented in [14], constituting then a novelty. It is complemented showing an overview of the plasticity process proposed. Finally, the conclusions.

2 Systems with Implicit Plasticity

We define a system with *implicit plasticity* as a system capable of detecting contextual information—where it is also considered the user and the hardware resources as part of the context—, taking advantage of it to adapt its own appearance to different contexts of use. In this sense this concept comprises both the *adaptivity* purpose as well as the *context-awareness* one. In accordance with Dix [9], *adaptivity* is “the modifiability of the UI by the system”, referring to an automatic modifiability and taking into account exclusively the user profile.

In our opinion, the concept of *implicit plasticity* can be considered equal to the *context awareness*—which is a decade old—, since the moment in which the properties related to the user as well as the restrictions in hardware resources are incorporated as properties that characterise the context. The goal is that all of them are taken into consideration in the adaptation process. In this way, our conception about context corresponds to the Schilit one in [17], who divides it in three categories: computing context, user context and physical context. In [1, 7] a detailed survey is offered.

The combination of varied contextual properties let us attain a more precise understanding of the current situation. However, apart from the treatment of the location and the user modelling, any other context parameter has been so widely studied [7]. Moreover, it has not been employed too much effort in exploring techniques that combine the user modelling with the context modelling, as is exposed in the next section. To develop an implicit plasticity engine and being able to adapt the system behaviour, it is necessary to know the environment that surrounds it, in order to characterise the context. Next, it must be established the method to model and to integrate such context in the system performance. The aspect of how to effectively use that information is still a challenging problem for developers of this type of systems.

It is worthy to remark that, despite the apparent correlation and proximity between the adaptation provided by the user modeling and the notion of context-awareness, the

explicit combination of both techniques is a relatively unexploited research area, as it is analyzed in [3]. In fact, to date there are few clear examples of developed applications that explicitly combine sensor-based context-aware techniques in conjunction with traditional user modeling techniques.

As user modeling-based applications that implicitly use context data as a part of their user models we can mention, for example, user's interests in [2], and user's tasks in [21]. Similarly, there are few examples of context-aware applications that incorporate the notion of user model to support their adaptive behavior. As a significant example we can mention the GUIDE system [8], that uses a user model in order to represent visitor's interests, preferences and the attractions that they have already visited.

Nevertheless, there exists a great amount of context-aware applications that work without the intervention of a user model. We can mention the applications within the field called memory prosthesis, that in the early nineties became an important application domain in the research field of context-aware computing, introducing the notion of augmenting human memory. Examples in this field can be consulted in [3].

To finish this analysis, we want to comment that some of the challenges raised by the incipient everyday computing [1] are to consider the techniques that have arisen from the approaches developed through user-modeling and context-aware computing. This term describes a paradigm of use in which constant interactions may occur between users and ubiquitous computing resources, implicitly or explicitly, based on continuously changing user contexts.

3 Our Proposal

3.1 Approach for the Implicit Plasticity Engine

Design. On the client side (the mobile device) our aim consists of developing an engine capable of adapting a specific UI in real time according to a set of conditions related to the context of use. These conditions that we will refer henceforth as *real time constraints*, should be represented and treated separately from the core functionality of the system, as non-functional properties that affect the UI. The aim is to maintain a self-representation (explicit modeling) of the part of system which we are interested in accessing, controlling and manipulating (the UI, together with the contextual factors that characterize it), in order to apply the necessary adjustments to the system's behavior, in a transparent and dynamic way. This is precisely what makes up the essence of a reflective system or architecture [12]. Therefore we use *reflection*⁴ to implement the implicit plasticity engine that we propose, due to its capacity of self-recognition and self-modification in runtime, extending and adapting its own behavior.

⁴ Under an object-oriented reflective architecture view a system is considered to be made up of two parts: the application core and the reflective part to be separated in thus called *base level* and *metalevel*, respectively.

Following this approach, the components related to the functionality of the application are represented in the base level of the reflective architecture (*logic layer* in the figure 1), and they are manipulated by the metalevel (*adaptive layer* in figure 1). Furthermore, the self-representation of the system is located in the metalevel, making directly explicit and accessible those properties that usually are not so. As these properties are normally common to diverse systems it is interesting to build these components as reusable. Lastly, the metalevel is also responsible for controlling changes in the different contextual factors. A third level (*context-aware layer* in figure 1) will be in charge of detecting the environment and notifying the changes to the meta-objects. Hence, following the principle of abstraction, in the base level the system works without any conception of interface which is relegated to the metalevel where is being continually re-modeled during run-time, reflecting the changes in the base system opportunistically.

Undoubtedly, the separation between the functional (*base level*) and the non-functional parts (*metalevel*) proposed by means of reflection prevents the non-functional properties from interfering with the basic functionality of the application, bringing flexibility in the development of the core of the application. However, the treatment of the non-functional properties is still being carried out as a whole, causing the same problems of code scattering and code tangling, although now located in the metalevel. To solve that and contribute to a major flexibility and reusability we propose to model the real time constraints, which constitute inherently crosscutting concepts, as *aspects*⁵ to influence in the metalevel, using AOP. Figure 1 resumes this idea. The notation used, which is based in UML, has been taken from [13].

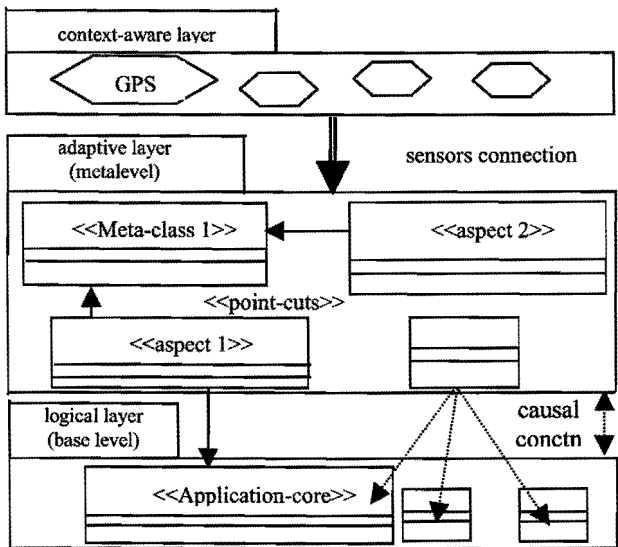


Fig. 1. Architecture for the implicit plasticity engine

⁵ Program unit to encapsulate crosscutting concerns.

Implementation Tools. To carry out the adaptive layer structure presented before, in which reflection and AOP techniques intervene, and after having analyzed diverse platform options, we have decided to adopt the J2ME (Java 2 Micro Edition) language, in combination with the *AspectJ* as an aspect oriented extension for Java. This platform guarantees the integration between both engines presented, as well as the dissemination of the implicit plasticity engine to the most wide range of mobile devices, including cellular phones due to the multi-platform character of Java. In this sense, although in the case of the example presented here it will not be feasible to use a cellular phone to support an tele-aid system, we are thinking in terms of a generic implicit plasticity engine. Another interesting alternative is using the .net platform. However, it is not clear its performance in other mobile devices but PDAs. Furthermore, the AOP solution offered is not so popular and supported as AspectJ.

In relation to that, one of the most important reasons why we have chosen Java is that the solution offered by *AspectJ* is the most extended and consolidated solution to AOP, as is proven in [22]. It constitutes the most simple and practical extension of AOP for Java and it has a lot of strong points. Moreover, the weak point of Java, which is reflection capacity, can be compensated by means of the *AspectJ* weaving possibilities. Finally, we would like to remark that as *AspectJ* generates pure Java byte code, applications using it will be able to be deployed in any mobile device that supports J2ME, which definitively are composed of the Java Virtual Machine. *AspectJ* uses the term *pointcut* (appeared in figure 1) to refer to those points in the code from which the code corresponding to the crosscutting concerns has been extracted in order to be injected afterwards. The entity *aspect* is in charge of intercepting them making possible reification and reflection.

3.2 Approach for the Explicit Plasticity Engine

Initial Development Framework. On the server side our aim consists of developing a plastic UIs development framework capable of producing the specific UI suitable in each case in a systematic way, avoiding having to manually develop as many UIs as there are possible contexts of use. We need a model-based framework. Let us go to outline a general description of the framework we have devised. These are the main characteristics:

- Composed of two sequential phases called *Abstract Rendering Process* (ARP) and *Concrete Rendering Process* (CRP) respectively, in which vary the set of models to take part. The first stage is in charge of obtaining the Abstract User Interface (AUI). The models that intervene are the next ones: Spatial Model (SM), Task Model (TM), Domain Model (DM), User Model (UM) and Dialog Model (DgM). The second stage manages the selection of the set of final interactors, which reside in the Presentation Model (PM), according to all the contextual information represented in the next models: UM, SM, Platform Model (PltM) and Contextual Model (CM), and also ruled by the DgM. More concretely, it is in charge of translating each abstract interface object in the AUI to a concrete interface object according with the current situation. As a result, this stage obtains the expected Concrete User Interface (CUI), resulting from the restrictions propaga-

tion. A complete description of all of these models as well as about both kinds of UIs can be consulted in [16].

- We propose to use model repositories. This allows each model to populate a common area with the specific concepts it is responsible for capturing. We use a model repository for each rendering process, making possible to share concepts between the models.
- Equally, we consider necessary to use two groups of mapping rules, one for phase, to manage the relations in each group.
- There also intervene some ergonomic heuristics, style guidelines and usability patterns [20] in the second phase to manage the transformation from Abstract Interaction Objects (AIO) to Concrete Interaction Objects (CIO), according to some environmental circumstances and to preserve usability.

This model corresponds to a *shared model approach* that allows models to share concepts as well as to be informed of any change produced in the UI being adapted, thereby providing the propagation mechanism we pursue. This framework arises from a deep analysis of the existing techniques based on models, presenting a solution to certain detected shortcomings, as it is shown in [16].

Figure 2 shows the scheme of our plastic UIs development framework, depicting all the relations among the models. It can be consulted [15] for more details in the relations and in the whole framework.

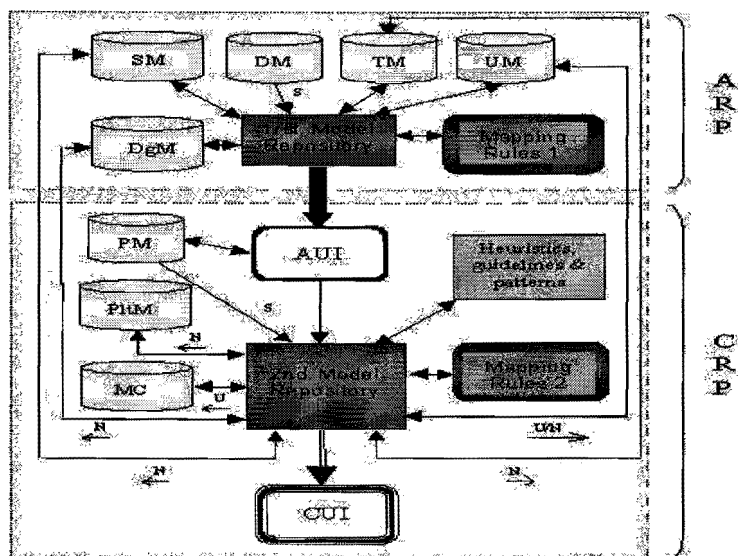


Fig. 2. Plastic UIs development framework (Explicit P.)

A semantic Approach to this Framework. This framework can be considered as a correct structured way to capture, store and manipulate multiple elements of the context of use. However, it is not flexible enough to consider multiple parameters for *universal design principle* [10]. The actual practicalities of arranging a systematic plastic UIs development framework which takes into account all of the aspects that take part in the context of use present some important problems related to the flexibility in contents. It means that it is pretty difficult to progressively take multiple parameters for universal design into account.

Here we propose incorporating a conceptual layer to locate an ontology of the domain of discourse. The conceptual level is therefore intended to enable domain experts to identify common concepts, relationships and attributes involved in any particular way in a universal design without limiting the set of parameters to be considered, instead allowing the incorporation of external information. Once the ontology is defined, the set of concepts and attributes defined can be instantiated for each context of use of a domain to obtain the suitable set of models. In short, the ontology defined will determine not only the models to intervene, but also the information they have to store in order to derive a *conceptual model*⁶ as flexible as possible, following the universal design principle.

This approach constitutes an improvement to the initial proposal in [14].

3.3 Overview of the Plasticity Process

Figure 3 shows the overview of the plasticity process that we want to transmit.

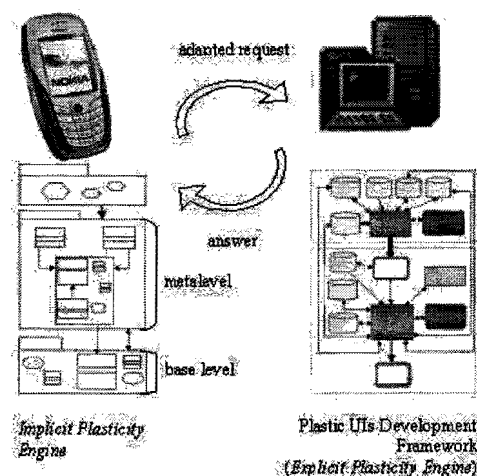


Fig. 3. Overview of the Plasticity Process

⁶ Formal, declarative and implementation-neutral description of the UI, that should be expressed by a modelling language.

The *implicit plasticity engine* is located in the client side. It is supported over a three-layered architecture as it has been presented. The *explicit plasticity engine* is located in the server side. It consists of a plastic UIs development framework, where all the models considered relevant in the configuration of the specific UIs intervene to produce specific UIs, as it has been presented. This one is not always active, but rather it acts upon client demand, producing a new UI suitable for the current circumstances. The implicit plasticity engine is in charge of launching the requests as much adjusted to the runtime conditions as possible, as well as of maintaining the connection to the servers. The request for the reconfiguration of the UI implies a process of updating of the involved models as required, thereby guaranteeing the propagation of contextual changes and the necessary feedback for the benefit of plasticity. That will establish an iterative and alternative process not established beforehand, and that depends on the circumstances, putting both engines in collaboration and guaranteeing the updating of the contextual situation on both sides.

4 Conclusions

Though the benefits of incorporating a context-sensitive and personalized component into a mobile application may be immense, the actual practicalities of doing so present some difficult design and implementation problems. A successful solution to the problem will mark a milestone in mobile computing.

This paper tackles this problem so noticeable. Firstly we identify and set bounds to two levels of operation in the plasticity process, to which a different subconcepts and different goals are assigned. Secondly we present our infrastructure to guarantee the plasticity of UIs consisting of a client/server-based architectural framework.

For the client side we present an architecture to support the implicit plasticity engine, in which we combine the reflection and de AOP techniques. On the one hand the reflection raises as a mechanism to solve the context-awareness, allowing designing for change. On the other hand, the AOP offers a mechanism to modularize and decouple in a optimum way the crosscutting concerns, obtaining so the desired orthogonality in its treatment. Definitely, we can assert that the structure we propose here for the adaptive layer satisfies a double purpose. Firstly it fulfils with the initial expectances -offering an adaptive behavior and UI- and secondly it is adjusted to the canons of flexibility, reusability and smartness we demand today to any software product. Therefore it constitutes a very valid option to model and manage the intrinsic concepts to the systems with implicit plasticity.

For the server side we present a model-based framework based on ontologies. We believe that the use of ontologies in a model-based framework is a decisive step towards the consolidation of the plastic UIs development framework we pursue. On the one hand, ontologies introduce a semantic component, and on the other hand they facilitate multi-disciplinarity and reusability based in the fact that changes carried out at any level are instantly propagated to subsequent levels.

References

1. Abowd, G.D., Mynatt, E.: Charting Past, Present and Future Research in Ubiquitous Computing. *ACM Transactions on Computer-Human Interaction*, Special issue on HCI in the new Millenium, Vol 7(1) (2000) 29-58
2. Ardissono, L. and Goy, A. Dynamic Generation of Adaptive Web Catalogs. *Proc. of Conference on Adaptive Hypermedia and Adaptive Web-base Systems (Trento 2000)*
3. Byun, H.E., Cheverst, K.: Exploiting User Models and Context-Awareness to Support Personal Daily Activities. *Proc. of User Modelling for Context-Aware Applications (2001)*
4. Calvary, G., Coutaz, J., Thevenin, D.: Supporting Context Changes for Plastic User Interfaces: a Process and a Mechanism. *Proceedings of IHM-HCI'2001 (2001)* 349-363
5. Calvary, G., et al.: Plasticity of User Interfaces: A Revised Reference Framework. *Proceedings of TAMODIA 2002 (2002)* 127-134
6. CAMELEON – Context Aware Modelling for Enabling and Leveraging Effective interaction. <http://giove.cnuce.cnr.it/comeleon.html> (2003)
7. Chen, G., Kotz, D.: A survey of context-aware mobile computing research. Technical Report TR2000-381, Computer Science Department, Dartmouth College (2000)
8. Cheverst, K., Davies, N., Mitchell, K., Smith, P.: Providing Tailored (Context-Aware) Information to City Visitors. *Proc. of Conference on Adaptive Hypermedia and Adaptive Web-based Systems (Trento, 2000)*
9. Dix, A., et al.: *Human-Computer Interaction*, 2nd ed., Prentice Hall Europe (1998)
10. Furtado, E., et al. An Ontology-Based Method for Universal Design of User Interfaces. *Workshop on Multiple User Interfaces over the Internet (2001)*
11. Kiczales, G., et al.: Aspect-oriented Programming, in M. Aksit and S. Matsuoka, ed., 11th ECOOP'97, Vol. 1241 of LNCS (1997) 220-242
12. Maes, P.: Concepts and Experiments in Computational Reflection. *Proc. of the 2nd OOPSLA'87 (1987)* 147-156
13. Pawlak, R., Duchien, L., Florin, G., Legond-Aubry, F., Seinturier, L., Martelli, L.: A UML Notation for Aspect-Oriented Software Design. *Proc. of AO Modelling with UML Workshop at the AOSD 2002 Conference (2002)*
14. Sendin, M.: Dichotomy in the Plasticity Process: Architectural Framework Proposal. *Advances in Pervasive Computing*. Austrian Computer Society. Doctoral Colloquium in Pervasive Computing (Linz 2004) 141-147
15. Sendin, M., Lorés, J.: Plastic User Interfaces: Designing for Change. *Proc. of the Making model-based UI design practical: usable and open methods and tools Workshop (IUI-CADUT'04) (Madeira, Portugal, 2004)*
16. Sendin, M., Lorés, J.: Hacia un Motor de Plasticidad Explícita Retroalimentado, in *Proceedings of the V Congreso en Interacción Persona-Ordenador (2004)*
17. Schilit, B., Adams, N., Want, R.: Context-aware Computing Applications. *Proc. of IEEE Workshop on Mobile Computing Systems and Applications*. IEEE Computer Society Press. (Santa Cruz, California, 1994) 95-90
18. Thevenin, D.: *Adaptation en Interaction Homme-Machine: Le cas de la Plasticité*. PHD Thesis, Joseph Fourier University, Grenoble (2001)
19. Thevenin, D., Coutaz, J.: Plasticity of User Interfaces: Framework and Research Agenda. *Proc. of Interact'99, Edinburgh (1999)* 110-117
20. Tidwell, J. UI Patterns and Techniques. <http://time-tripper.com/uipatterns> (2002)
21. Vassileva, J.: Ensuring a Task-based Individualized Interface for Hypermedia Information Retrieval through Use Modeling, in Maybury M.T. (ed.) *Intelligent Multimedia Information Retrieval*, MIT Press (London, 1997)
22. Xerox PARC.: AspectJ home page. Web (2002)

Adaptive Hypermedia in the AdaptWeb Environment

Mariusa Warpechowski¹, José Palazzo Moreira de Oliveira¹, Maria Aparecida Martins Souto¹, Daniela Leal Musa¹, Maria Angélica de O. Camargo Brunetto², Veronice de Freitas², Mário Lemes Proença Jr.²

¹ Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Av Bento Gonçalves, 9500, Porto Alegre, RS, Brazil

² Departamento de Computação – Universidade Estadual de Londrina (UEL), Campus
Universitário, Londrina, PR, Brazil

Abstract. This paper describes how the adaptive environment for Web-based learning – AdaptWeb[®] – provides educational content personalized according to a specific university program and to a particular student profile. Hypermedia teaching applications can explore different educational strategies and tactics, including guided navigation, hierarchical contents presentation, examples, exercises, and so on. Also the teaching task can be optimized if the teacher prepares a content material that can be shared for different target public. AdaptWeb[®] was designed concerning these aspects and allows the generation of different presentations from a given discipline based on the student model. When the student follows a particular presentation, the system AdaptWeb[®] verifies not only the student's program but also the particular preferences and technological environment. The student model is stored and updated using MySQL database system. The specific presentation is generated dynamically by analyzing the current log file of the student and using the appropriate XML document according to his student program.

1 Introduction

The need for distance learning/teaching made distance education environments (DEE) via Web to spread out all over the world, thus needing constant updating and refinement. The diversity of student's preferences and abilities requires adaptive techniques to support DEE over the Web [1]. Thus, there is a need for the creation of tools that allow authors to apply adaptive techniques when making content available for different user's profiles, but instead of creating a complete system for each profile, a specific guide creating a special university programs associated to each profile is created. AdaptWeb[®] environment [2], described in this paper, was developed with the aim of fulfilling this need. The beta version, implemented in PHP and MySQL is available at SourceForge.

According to Brusilovski [3], an adaptive system collects data for the user's model from different sources, which can implicitly include the observation of the user's interaction and explicitly the request of data input made directly by the user. The user's model is used to adapt the environment in order that it suits the different user's needs in a single context. In AdaptWeb[®], adaptation is based on the student's model, which

is formed by information about the student's background (university program), his/her preference for navigation and navigation history. AdaptWeb[®] is composed of four modules that have different and complementary functions, from the development of multiple presentations of a content for a discipline and interface adaptation based on the student's model.

The present paper is organized as follows: section 2 presents related works, section 3 presents the environment, sections 4, 5, 6 and 7 describe each module. Finally, section 8 presents some conclusions.

2 Related Work

The Adaptation technologies are defined for Brusilowvski [1] as different forms of adaptation in the educational system based on the Web. Two technologies distinguish among others: the adaptive presentation and adaptive navigation. The goal of the adaptive navigation is to support the student in the hyperspace orientation and navigation by changing the appearance of visible links, therefore the adaptive presentation technology adapts the content of a hypermedia page to the user's goals, knowledge and other information stored in the user model [1].

The Systems using adaptive presentation include Hypadapter [4], which adapts through Providing Explanation Variants and Reordering Information, and AHA [5] through fragments of a page that can be conditionally included or excluded, and links that can be conditionally hidden or presented in a different colour. The systems like ELM-ART [6], ISIS-Tutor [7], and AHA supports adaptive navigation. The ELM-ART use link annotation, ISIS-Tutor uses link annotation and link removal, and the AHA uses link annotation and link hiding.

The Adapt Web supports adaptive presentation, through Conditional Inclusion of Fragments, where concepts are release after the prerequisite studies, and Reordering Information, where the student can choose the sequence and access examples, exercises or complementary material. The system also supports adaptive navigation that uses: links annotation, link disabling, and link removal and map adaptation. The adaptation on Adapt Web will be detail in the following sections.

3 The Environment

AdaptWeb[®] environment offers the following adaptable options: navigation (free or tutorial), presentation (studied, not studied and under study concepts, available and unavailable contents) and content (concepts structure put available according to the student's university program). The objective of this environment is to enable the adequacy of instructional contents to students from different university programs, with distinct personal preferences, according to the student's model [8]. The architecture is composed of 4 modules as Fig. 1 shows: the Authorship module (1), the Storage module in XML (2), the Instructional Content Adaptation module (3) and the Adaptive Interface module (4).

The author registers the concepts structure organized and for each concept the author informs which university programs will be available, the pre-requirements and material related. The system creates XML files, which are filtered to generate the instruc-

tional content adapted to the student's university program. The content is displayed to the student in an interface that helps his/her navigation through content.

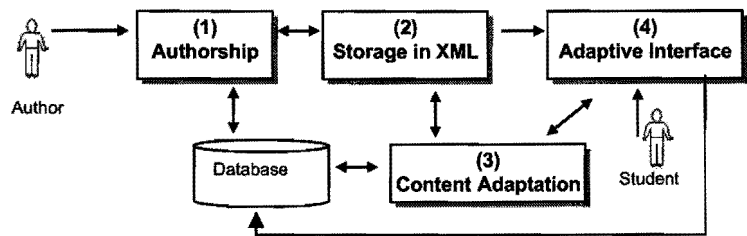


Fig. 1. AdaptWeb System Architecture

4 Authorship – Structuring and Organizing Content

The Authorship module is responsible for structuring and organizing instructional

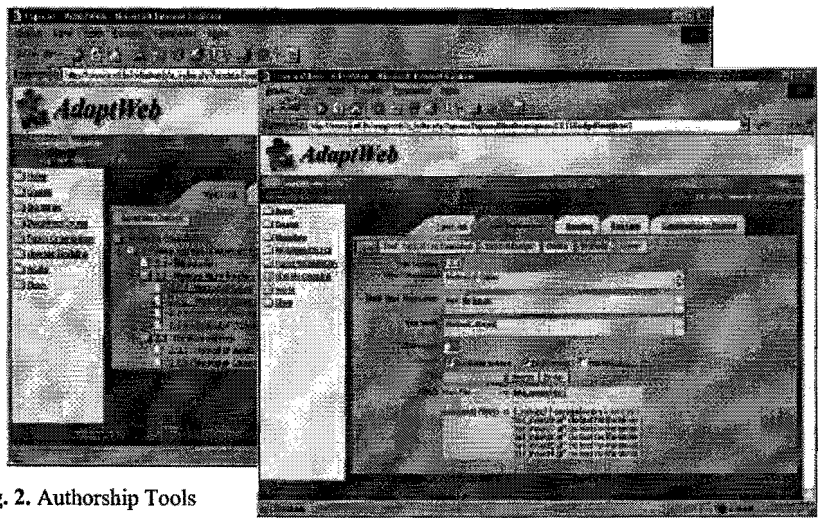


Fig. 2. Authorship Tools

content that will be available to the user. In this module the content of some subject is organized through concepts, hierarchically, in a single structure adapted to each different university program. The instructional content is composed of educational resources as: Concept, Exercise, Example and Complimentary Material.

The author uses an authorship tool, where he/she registers all concepts and materials related to each concept. For each concept registered, the author informs description, keywords, pre-requirements, for which university programs he/she wants to make some concept available and the file referring to the concept. In this phase, the author can also specify the exercises, examples and complimentary material, which can also

be adapted to the student's university program. Figure 2 shows the authorship module screens.

5 Storage of the Content Structure in XML

AdaptWeb[®] generates XML (*Extensible Markup Language*) files through the Storage module in order to store the data structure in a memory originated in the authorship phase. For the creation of XML files, two DTD (*Document Type Definition*) were defined. A DTD depicts the hierarchical structure of concepts and the other DTD describes the specific content of each concept.

An algorithm to store the content structured was defined based on DTD. This algorithm creates an XML file for each subject with its respective concepts structure and the features of each concept and XML file to each concept, that include tags for concepts, examples, exercises and extra material in its body. Storing files in XML format makes possible to structure data in a hierarchical way, because there is always a single XML file with the structure of concepts of the subject and as many XML files with content as the concepts defined. The XML files generation is validated through a *parser* that scans the documents.

6 Content Adaptation based on the Student's Model

The Content Adaptation module is responsible for the selection of instructional contents adequate to its model, considering the subject and the student's university program. For that end, it selects contents from an XML document, which contains the reference for all kinds of contents associated to a concept. The content adaptation module takes as input an XML file that is accessible through the DOM API, making its content available in the form of a tree, which will be adapted according to the student's university program features, generating an instance of XML file.

In order to select the educational content to be presented to the student, filters are applied to XML files of each subject. The first filter, applied to XML file of the concepts structure, generates a new tree of this document containing only the concepts that make part of the university program for the student in question. Yet, during the application of this filter a log table of the student is checked. It contains information on the navigation through the subjects' concepts. After information from the log it is possible to mark in the XML document which concepts the student accessed and which is the current concept, that is, the last concept visited. The second filter is applied to XML files that store the educational content, during the student's navigation through concepts (concept (theory), examples, exercises and complimentary material). According to the chosen concept and to the student's university program, the filter is applied over the respective XML content file and the presentation is dynamically generated.

7 Adaptive Interface

In Adaptive Interface module the adaptation occurs in the links that are available for the user, according to his/her model. The Adaptive Interface module checks which concepts are related to a certain group of users (university program) in the concepts structure XML. Figure 4 shows the Student's environment. The navigation history provides information to the Adaptive Interface module about the pre-requirements the author defined in the Authorship module. If the concept of pre-requirement was studied by the user, then the concept that depends on it can be enabled. A concept can assume three categories: 1) "studied", it means that the student has already accessed the concept; 2) "under study", it refers to the concept that is being accessed, also named current; 3) "not studied", it can be in this category for two reasons, (a) the concept was not enabled because its pre-requirements were not studied yet; or (b) the concept was not studied yet, but is enabled after pre-requirements were studied.

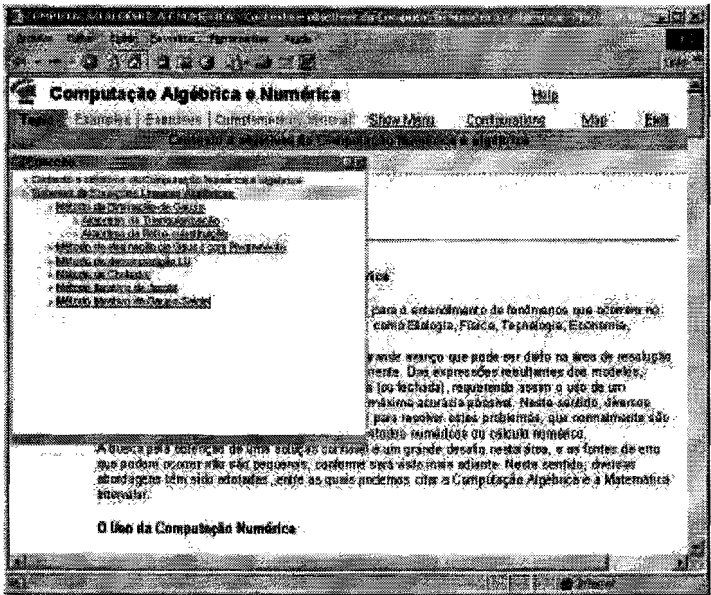


Fig. 4. Student's environment.

According to the navigation mode, the adaptation module can work in two ways, in the tutorial mode, pre-requirement criteria among concepts determine the student's navigation, and navigation adaptation is based on the register of concepts studied: at each new access of the same student, colors of the menu links are restored and in the free mode, the student can study any concept available in the navigation menu. Colors used follow the usability rules by Nielsen [9], that is, we adopted blue for links that were not accessed, and violet for the accessed ones. For the implementation of the navigational adaptability, methods of links disabling, links annotation and link removal [10] were used.

8 Conclusions

This paper presented AdaptWeb[®], which makes easier the author task of offering different content according to the student's profile. The pre-authorship systematic and the authorship tool allow the author to organize the content in a way that it adapts to the student, without losing time and effort.

The environment provides adaptability of the educational content and the interface based on the student's model. Features of the student's model considered are adequate to adapt content and interface to different users of the system. User's formation, that is, the university program in which he/she takes part makes possible that the system makes available only the content that is adequate to the university program, speeding up the student's learning. The navigational preference, tutorial or free, allows the student to navigate according to his/her preference, following the steps the author determined or navigating freely, stepping up some topics already known. Analyzing the user's navigational history, the environment indicates where the student stopped and the point where from he can continue. The present implementation offers only a basic set of adaptability; the most important characteristic of this environment is the ability to include the student's cognitive style characteristics in the learning content adaptation. This last point has been worked in two Ph.D. dissertations and is now being deployed to the software environment. Thus, we have an environment where content presented to the student is according to his/her objectives and preferences, providing a more efficient learning.

References

1. Brusilovsky, P.: Adaptive Educational Systems on the WWW: A review of available technologies. In Proceedings of the Workshop "WWW-Based Tutoring" at the 4th ITS. 1998.
2. AdaptWeb – Homepage and download at SourceForge: <http://adaptweb.sourceforge.net>, a running demo is available at: <http://adaptweb.homelinux.org>
3. Brusilovski, P. and Maybury, M. T.: From Adaptive Hypermedia To The Adaptive Web. In. Communications of the ACM 45 (5), Special Issue on the Adaptive Web, 31-33. 2002
4. Hohl, H., et al.: HYPADAPTER: An Adaptive Hypertext System for Exploratory Learning and Programming, User Modeling and User-Adapted Interaction, 6(2-3), 131-155, 1996.
5. De Bra, P., Clavi, L.: Creating Adaptive Hyperdocuments for and on the Web in Proceedings of the AACE WebNet Conference, Toronto, 149-155, 1997.
6. Brusilovsky, P., Schwarz, E. and Weber.: Elm-art: An Intelligent Tutoring System on World Wide Web. In ITS, Montreal, LNCS 1086, 261-269, 1996.
7. Brusilovsky, P., Pesin, L.: Adaptive Navigation Support in Educational Hypermedia: An Evaluation of the ISIS-Tutor. In Journal of Computing and Information Technology, 6(1), 27-38, 1998.
8. Freitas, V., et al.: AdaptWeb: an Adaptive Web-based Courseware. ICTE : International Conf. on Information and Communication Technologies in Education. Badajoz, Espanha. 2002
9. Nielsen, J. Designing Web Usability: The Practice of Simplicity, Indianapolis: New Riders Publishing. 2000
10. De Bra, P. Brusilovski, P., Houben, G.: Adaptive Hypermedia: From Systems to Framework, ACM Computing Surveys 31, 4. 1999.

WORKSHOP 2:
Individual Differences in Adaptive Hypermedia

Workshop Co-Chairs:

George D. Magoulas
Sherry Y. Chen

Preface

The Workshop on Individual Differences in Adaptive Hypermedia is part of the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-based Systems that was held from August 23 to August 26, 2004, at The Eindhoven University of Technology, The Netherlands.

The Workshop explores how to embrace the various dimensions of individual differences into adaptive hypermedia, and investigates the impacts of individual differences on the design, implementation and use of adaptive hypermedia systems.

Individuals differ in traits such as skills, aptitudes and preferences for processing information, constructing meaning from information and applying it to real-world situations. However, existing applications mainly consider users' preferences based on collecting explicit or implicit information, and emphasise on prior knowledge. As a result, it is still not very clear whether adaptive hypermedia systems can accommodate individual differences effectively, in terms of providing individualised navigation support, delivering personalised content, adapting the presentation or the layout to the needs of the user.

The contributions that are presented here cover various dimensions of individual differences, such as the level of knowledge, spatial abilities, learning styles, cognitive styles, accessibility issues and seek to provide answers to the following questions:

- How adaptive hypermedia can improve accessibilities by providing multi modalities that satisfy users with special needs?
- What design guidelines should be established for development, and what criteria are needed for evaluating adaptive hypermedia that can accommodate individual differences?
- How different dimensions of individual differences can be combined in an adaptive hypermedia system?
- What type of information is needed from user profiles to identify the effects of individual differences on user's preferences?
- What kinds of ontologies are needed for representing individual differences dimensions in the user model and the personalisation engine of adaptive hypermedia systems?
- What are the relationships between individual differences and features of adaptive hypermedia systems?

We hope that the Workshop will contribute to the global research in Adaptive Hypermedia by comprehensively reviewing state-of-the-art adaptive hypermedia approaches that accommodate individual differences, will help integrating individual differences theory into adaptive hypermedia applications, and will give some insight into analytical and architectural aspects of adaptive hypermedia that exploit individual differences for personalisation.

August, 2004
George Magoulas and Sherry Chen

Workshop Organisation

George D. Magoulas, *Birkbeck College, University of London, UK*
Sherry Y. Chen, *Brunel University, UK*

Program Committee

Ann Blandford, *UCL Interaction Centre, University College London, UK*
Peter Brusilovsky, *University of Pittsburgh, USA*
Nigel Ford, *University of Sheffield, UK*
Anthony Jameson, *DFKI-German Research Center for Artificial Intelligence, Germany*
Judy Kay, *University of Sydney, Australia*
Robert Macredie, *Brunel University, UK*
Kyparissia Papanikolaou, *University of Athens, Greece*
Alexandra Poulovassilis, *London Knowledge Lab, University of London, UK*
Diane Sonnenwald, *Göteborg University and University College of Borås, Sweden*
Marcus Specht, *Fraunhofer Institute for Applied Information Technology, Germany*

Additional Reviewers

Sara de Freitas, *London Knowledge Lab, University of London, UK*
George Ghinea, *Brunel University, UK*
Mark Levene, *Birkbeck College, University of London, UK*
Lionel Sacks, *University College London, UK*
Demetrios Sampson, *University of Piraeus, Greece*

Accommodating learning style characteristics in Adaptive Educational Hypermedia Systems

K.A. Papanikolaou^{1,2} and M. Grigoriadou²

¹ Technology Education and Digital Systems, University of Piraeus, Greece

² Department of Informatics & Telecommunications, University of Athens,
Panepistimiopolis, GR-15784 Athens, Greece
{spap,gregor}@di.uoa.gr

Abstract. In this paper we build on research reported in the areas of Adaptive Educational Hypermedia and learning styles in order to deal with critical issues influencing the design of adaptation based on the learning style information. In more detail we concentrate on: (i) the different learning style categorizations that have been or could be used for modelling learners' learning style in the context of an Adaptive Educational Hypermedia System and the way these could guide the design of adaptation, (ii) the adaptation technologies that could better serve learners with different learning styles, (iii) the dynamic adaptation of the system and the diagnosis process including the identification of specific measures of learners' observable behaviour which are indicative of learners' learning style preferences.

1 Introduction

As learning styles are a significant factor contributing in learner progress, a challenging research goal is to attempt to represent specific characteristics of learners' learning style within Adaptive Educational Hypermedia Systems (AEHS). Taking into account that many different classifications of learning styles have been proposed in the educational psychology literature, this is a demanding task motivated by the expected learning benefits.

Important decisions underlying the incorporation of learning style characteristics in AEHS demand the synergy of computer science and instructional science, such as: (i) the selection of appropriate categorizations, which are appropriate for the task of adaptation, (ii) the design of adaptation, including the selection of appropriate adaptation technologies for different learning style categorizations and of appropriate techniques for their implementation, (iii) the design of the knowledge representation of such a system in terms of the domain and the learner model, (iv) the development of intelligent techniques for the dynamic adaptation of the system and the diagnosis process of learners' learning style including also the selection of specific measurements of learners' observable behaviour, which are considered indicative of learners' learning style and studying attitude.

The research goal of accommodating learning styles in AEHS design could also be combined with the development of meta-adaptive hypermedia systems capable of selecting the most appropriate adaptation technology following the individual characteristics of the current users and context (Brusilovsky, 2003). To this end, an AEHS should have a number of different adaptation technologies at its disposal and be aware about the limits of applicability of each technology. In this context learning style information can considerably contribute to the decision of the appropriate adaptation technologies for learners with particular profiles, as specific categorizations of learning styles seem to match better with specific adaptation technologies.

In this paper we build on research reported in the literature about different approaches that have been adopted for the design of adaptation based on the learning style information, in order to deal with critical issues for the development of an AEHS based on this information. In more detail we investigate: (i) the different learning style categorizations that have been or could be used for modelling learners' learning style in the context of an Adaptive Educational Hypermedia System and the way these could guide the design of adaptation, (ii) the adaptation technologies that could better serve learners with different learning styles, and (iii) the dynamic adaptation of the system and the diagnosis process including the identification of specific measures of learners' observable behaviour which are indicative of learners' learning style preferences.

2 Learning style information in Adaptive Educational Hypermedia

Designing adaptation based on the learning style information builds on hypotheses about the relationship of learning behaviour with learning style. Such hypotheses are necessary for modelling the learners' learning style in the context of an AEHS. Valuable resource in this context is research conducted in the area of educational psychology about learning styles and the way this characteristic influences learners' behaviour and preferences. A variety of learning style categorizations has been proposed which attempt to associate specific characteristics to different categories of learners and propose instruments and methods for assessing learning style (Riding and Rayner, 1998). Such categorizations could provide the necessary theoretical background for designing the adaptive behaviour of an educational system and guide decisions about what the system should offer to learners with different styles and how to do it.

The last years several AEHS reported in the literature use learning style information as a source of adaptation (see Table II). Several of them build on a theoretical background inspired from the learning style research. The objective of this section is to investigate the way different categorizations of learning styles could support the design of adaptation in terms of specific adaptation technologies. To this end we investigate: (i) the learning style categories that have been or could be exploited in AEHS and the way several of them have been used for modelling learners' learning style, and (ii) the implications that different learning style categorizations have on the design of different adaptation technologies.

Modelling the learning style information. Sadler-Smith (1997) identified four broad categories of 'learning style' in an attempt to acknowledge and accommodate the range of aspects of individual differences referred in the educational psychology literature in an holistic way: (i) 'cognitive personality elements' such as field dependence and field independence (Witkin et al., 1977), (ii) 'information-processing style' such as the experiential learning cycle (Kolb, 1984) and the associated leaning styles (converger, diverger, accommodator, assimilator), or the related learning styles suggested by Honey & Mumford (1992), activist, reflector, theorist, pragmatist, (iii) 'approaches to studying' such as deep approach, surface approach, strategic approach, lack of direction, academic self-confidence (Entwistle & Tait, 1994), (iv) 'instructional (i.e. learning) preferences' defined as an individual's propensity to choose or express a liking for a particular instructional technique or combination of techniques, such as dependent learners, collaborative learners, independent learners suggested by Riechmann & Grasha (1974). In this paper we use the term 'learning style' as a representative one for all the aforementioned categories.

In an attempt to organise the different approaches adopted in several AEHS reported in the literature, we identified: (i) systems that use the learning style information in order to design the content of instruction, and (ii) systems that use the learning style information in order to adapt to the learners' 'form' of cognitive activity (i.e. thinking, perceiving, remembering). The first class of systems usually adopt categorizations of learning styles that belong to the 'information-processing style' or 'instructional (i.e. learning) preferences' categories, while systems of the second class adopt categorizations of learning styles that belong to the 'cognitive personality elements' category.

In more detail, the adaptive behaviour of AEHS that belong to the first category concentrate on the type and usually the sequencing of material they offer based on a framework proposed by the authors (ACE, Arthur, MANIC) or based on research studies (Honey & Mumford, 1992), (Felder and Silverman, 1988) about the type of instructional material that learners with different learning style prefer (INSPIRE, CS383).

The systems of the second category, which are developed based on learner's cognitive style, concentrate on the 'form' of cognitive activity (i.e., thinking, perceiving, remembering) that learners usually adopt (Triantafillou et al., 2003; Bajraktarevic et al., 2003). For example, AES-CS (Triantafillou et al., 2003) uses the Field dependence/independence (FD/FI) styles [14]. AES-CS adopts several instructional strategies that accommodate learners' learning style in relation with: the approaches (global versus analytical approach), the control options (program control versus learner control), the contextual organizers (advance organizer, post organizer), the study instructions (provide minimum or maximum instructions), the feedback, and the lesson structure. Also, Bajraktarevic et al. (2003) use the Holist/Serialist learning styles proposed by Pask (1976) which is aligned with the Wholist/Analytics dimension and with the Global/Sequential categorisation (Felder and Silverman, 1988). Following the adopted approach, the system provides learners with different linking structures of the content tailored to their learning style.

Implications for Adaptation Design. The objective of this sub-section is to investigate the way different categorizations of learning styles that focus on different

characteristics of learners could support the design of adaptation in terms of specific adaptation technologies.

Adaptive presentation & curriculum sequencing. Adaptive presentation and curriculum sequencing technologies aim at tailoring the educational content to learners' learning style (adapt the content or its sequencing). These adaptation technologies could better serve learning style categorisations that deal with learners' preferences of instructional material or instructional strategies, such as those that belong at the 'information-processing style' or 'instructional (i.e. learning) preferences' categories. Representative examples of this approach are the systems Arthur, CS383, ACE, and INSPIRE. Arthur and CS383 use multiple types of resources differing in the media they utilize, whilst ACE and INSPIRE adapt the sequencing of different types of resources to different learning style categories following a variety of instructional strategies. In the first case, the alternative styles of instruction that are adopted for learners with different learning style demand the development of multiple types of educational material using different media for each particular section of the course. In the second case multiple types of resources are reused following a different sequencing based on the learner's learning style. This is an alternative to the commonly used approach of rewriting the same content for each learning style category (McLoughlin, 1999).

Adaptive navigation support. The goal of the adaptive navigation support technology is to support the learners in hyperspace orientation and navigation by changing the appearance of visible links. In this context the learning style information could serve as a valuable resource about learners' navigation "habits" and needs. Thus, the design of this technology could be mainly supported by research in the area of learning style categorizations that belong to the 'cognitive personality elements' and deal with the structure and organisation of the contents of instruction, such as the FD/FI dimensions and wholist-analytic dimensions. AES-CS is a representative AEHS that uses the learning style information in order to decide which navigational tools and aids are appropriate in order to help learners organize the structure of the knowledge domain and move accordingly within.

Adaptive collaboration support. Learning style information can also be used as the basis for the construction of groups to support collaborative learning. In the context of AEHS, the goal of the adaptive collaboration support technology (Brusilovsky, 1998) is to use system's knowledge about different users (stored in user models) to form a matching collaborating group. Thus, an interesting approach would be to use the learners' learning style information for organizing learners in groups as this characteristic is considered to influence social interaction. Thus, the design of the adaptive collaboration technology could be mainly supported by categorisations that deal with the social dimension of learners. For example, studies have identified a number of relationships between FD/FI dimension and learning, including the ability to learn from social environments (Witkin et al., 1977). Thus, FI individuals tend to enjoy individualised learning, while FD ones cooperative learning. Also, following Honey and Mumford (1992), groups with full range of learning styles in terms of Activists, Reflectors, Theorists and Pragmatists, exhibit better performance compared to randomly constituted groups.

Moreover, different learning style categorizations may assist the design of more than one adaptation technologies such as the verbal-imagery dimension. This learning style dimension interacts with mode of presentation of information (for example textual/verbal or diagrammatic/pictorial modes) and thus it may assist the design of the instructional material in the context of the adaptation presentation technology as well as the design of navigational aids in terms of the adaptive navigational support technology. Although experimental results are promising (see next section) more research has to be conducted in order to learn more about the relationships between learning styles, learning behavior in terms of observable patterns of learners' activity and possible adaptation approaches.

Open Issues. Although several learning styles categorizations have been exploited in AEHS, there are many more that have not been considered yet such as those that belong in the category of 'approaches to studying'. What is important in exploiting different learning style categorizations in AEHS is their potential to support and enhance adaptation providing appropriate guidance for AEHS developers. Thus, the wide range of learning style categorizations should be investigated through the ways each categorization could assist the design of the different adaptation technologies or inspire the design of new ones. This research goal has two different values both for the educational psychology area to evaluate the effectiveness and the validity of matching instructional methods to learners' styles and preferences in e-learning, and the adaptive educational hypermedia area to improve the effectiveness and efficiency of adaptation.

3 Evaluating the benefits from designing adaptation based on learning styles

Although several AEH systems that use learning style as a source for adaptation have been reported in the literature, just a few empirical studies (usually small scale studies conducted in experimental conditions) have been conducted that prove the effectiveness of the adopted approaches. The goals of such studies concentrate on the effectiveness and/or efficiency of adaptation, which are measured through learners' performance, learning time, navigation patterns, learners' subjective estimation. Different dimensions that are considered in these studies are: (i) the relationship between matching and mismatching instructional approaches with learners' learning style (Ford and Chen, 2001; Bajraktarevic et al., 2003) (ii) the learning performance and learning time of learners with different learning style in matched sessions (Triantafillou et al., 2003); (iii) the navigation patterns of learners with different profiles in matched sessions (Papanikolaou et al., 2003).

Ford and Chen (2001) investigated if the matching of instructional presentation strategies and learners' learning style is linked with improved learning performance. They report that learners of the FD/FI styles who learned in matched conditions scored significantly higher in tests measuring their conceptual knowledge but not in performing practical tasks. Following the authors, these results provide evidence about the learning benefits coming from matching learners' learning style with instructional presentation strategies and indicate the need to take into account

qualitative characteristics of expected learning outcomes such as learning, recall and application of conceptual knowledge, in designing adaptation. Triantafillou et al. (2003) conducted a small group evaluation in order to measure the effectiveness and efficiency of the instructional approaches adopted in AEC-CS for FD/FI learners. They found that all learners' performance was increased after instruction in matched conditions. In more detail, they found that the FI learners had better results than the FD ones, although FD learners were improved more than the FI ones. Furthermore, as learners spent less than an hour to complete the courseware (which was designed to correspond to a typical lecture hour), the adopted approach was considered efficient. In this study learners reported their satisfaction from the initial adaptation as well as from the fact that the system was completely controllable by them. In another study reported in (Papanikolaou et al., 2003), the authors analyzed learners' studying behaviour (time spent and hits on resources) and navigation traces by the different learning style categories proposed by (Honey and Mumford, 1992). The main aim of this study was to provide evidence about the way learners that belong to different learning style categories select and use educational resources that are considered beneficial for their styles in INSPIRE. Although this was a pilot study, the results were encouraging, confirming the initial hypotheses on which the presentation and sequencing of resources was based. Lastly the main aim of the study reported in (Bajraktarevic et al., 2003) was to evaluate the effectiveness and the efficiency of the adopted adaptation approach. Effectiveness was measured through learners' performance in matched and mismatched learning-style sessions (Holist/Serialist learning styles), whilst efficiency through learners' browsing time in matched and mismatched learning-style sessions. Learners' performance was significantly higher in matched sessions for all learners, whilst there was not significant difference between browsing times for the matched / mismatched groups.

4 Diagnosis of learning style: critical issues influencing adaptation

Vermunt (1996) conceptualises learning styles as consistent patterns of learning activities that are systematically linked to learning beliefs and motivational orientations. Thus, learning styles are not taken to be invariable (at least many of the proposed categorisations), as they may be influenced by the particularities of the learning context and its demands. Along this line, in the context of AEHS, a critical issue for recognising changes in learners' needs and preferences is to determine measures of learners' observable behaviour which are indicative of learners' learning style preferences. Thus, incorporating the learning style information in the context of AEHS requires, apart from a theoretical background, a qualitative analysis (categorization) of learners' steps and/or selections (features/tools of the system that they access/use) as they interact with the system. This information is also valuable in order to study the extent to which the hypotheses about learners' learning style preferences, match their learning behaviour as it is depicted through their actual navigation through the interaction.

Student diagnosis is the process of inferring students' internal characteristics from their observable behavior (VanLehn, 1988). An AEHS, due to the restricted

communication channel, is only able to directly obtain raw measurements, by monitoring the interaction with the learner, aiming to identify learners' changing needs and maintain the current state of the learner. Thus, critical issues that should be considered in designing the diagnosis process of learners' learning style are: (i) the initialisation of the learner model, (ii) the selection of appropriate measures to serve as indicators of learners learning style preferences, and (iii) the qualitative analysis of learners' observable behaviour that could support the dynamic adaptation of the system during the interaction.

In this context diagnosis should exploit the two methods usually used for assessing learners' learning preferences (Riding and Rayner, 1998): self-report measures through questionnaires, and observed behaviour choices. Especially the first approach is usually adopted for the *initialisation of the learner model*, whilst the second one for the dynamic adaptation of the system through the interaction. Following the first approach, several systems use specially designed psychological tests designed for particular learning style categorisations (INSPIRE, AES-CS), whilst others use interviews in order to let the learners decide on specific aspects of their learning style preferences (ACE). During the interaction, several systems allow the learners to directly manipulate their learning style expressing their own point of view about themselves and consequently about system adaptation (INSPIRE, AES-CS).

Through the interaction with the system, learner's observable behaviour is, in many cases, the basis for the diagnosis of certain characteristics of the learner such as his/her preferences of the learning material. In such cases, the dynamic adaptation of the system is based on real data coming from learners' interaction with the system. For example, in ACE, the dynamic adaptation of the instructional strategy is based on information coming from monitoring learner's requests on learning materials, as well as on the success of the currently used strategy. The latter is mainly determined by learner's performance in the final tests; repeated occurrences of high performance raise the preference value of a strategy until a threshold is reached. Also, Arthur dynamically adapts the instructional style according to learner's performance in the tests s/he submits. For example, in case the learner scores 70% in a quiz of a concept, then s/he will be provided with material of alternative instructional style; otherwise, the instructional style currently used is supposed to match the learner's learning style. Lastly, MANIC uses machine learning techniques in order to identify learners' preferences by observing his/her interactions with the system.

The selection of measures on which dynamic adaptation is based, is a significant factor influencing its effectiveness. For example, is learners' performance on tests or time spent on educational resources, adequate measures for learners' changing learning style preferences during the interaction? What about the individual characteristics of the learning style categorisation adopted for modelling learners' style, or the hypotheses on which system adaptation is based about learners' style? For example in case of the FD/FI categorization the way learners navigate is more appropriate as an indicator of their style than the specific type of material they select. On the other side this may be a valuable information about categorizations such as Verbalisers / Imagers or Activists / Theorists / Pragmatists / Reflectors. Thus, a critical issue in designing dynamic adaptation based on learners' observable behaviour is to identify which learners' actions are indicators of their style, and

should be considered in assessing their changing needs and preferences during interaction.

To this end, valuable resources could be studies reported in the literature investigating which *measures of learners' observable behaviour* are indicative of their learning style preferences and learning behaviour. Indicators that have been investigated for several learning style categorizations are: (i) navigational indicators (number of hits on educational resources, preferable format of presentation, navigation pattern); (ii) temporal indicators (time spent in different types of educational resources proposed); (iii) performance indicators (total learner attempts on exercises, assessment tests) (Reed et al., 2000; Lu et al. 2003; Souto et al., 2002; Papanikolaou et al, 2003). This is a promising research direction which may help us develop deeper knowledge of the complex interactions between learners and educational content and further inspire new approaches in the design of AEHS.

4 Conclusions

Especially in a web-based educational system, where the variety of learners taking the same course is much greater, a challenging goal in the design, development and delivery of learning could be the accommodation of learners' individual differences in terms of their learning styles. Towards this end, critical issues on which research in AEHS should focus are: (i) the design of adaptation based on the learning style information (what the system should offer to learners with different styles and how to do it in terms of deciding which adaptation technologies could better serve the aims of the adaptation), (ii) the selection of appropriate measures of learners observable behaviour which could serve as indicators of learners learning style preferences, (iii) the qualitative analysis of these observable measures that could support the dynamic adaptation of the system during the interaction.

To the above research goals valuable resources are the different categorizations of learning styles proposed in the area of educational psychology. Such information may: (i) assist in the design of AEHS which accommodate learners' styles and preferences; (ii) contribute to the enhancement of the pedagogical perspective of such systems; (iii) assist the evaluation of the effectiveness and efficiency of adaptation; (iv) provide directions for future research into the validity of matching instructional methods to learners' styles and the effectiveness of adaptation.

References

1. Brusilovsky, P.: Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology* 34 (4) (2003) 487-497
2. Brusilovsky, P.: Adaptive Educational Systems on the World-Wide-Web: A Review of Available Technologies. In *Proceedings of Workshop "WWW-Based Tutoring" at 4th International Conference on Intelligent Tutoring Systems (ITS'98)*, San Antonio, 1998

3. Bajraktarevic, N., Hall, W., Fullick, P: Incorporating learning styles in hypermedia environment: Empirical evaluation. In: Proceedings of AH2003: Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, Budapest, Hungary, 2003 (41-52)
4. Carver, C.A., Howard, R.A., Lane, W.D.: Enhancing student learning through hypermedia courseware and incorporation of student learning styles. *IEEE Transactions on Education* 42 (1) (1999) 33-38
5. Ford, N. and Chen, S.: Matching/mismatching revisited: an empirical study of learning and teaching styles. *British Journal of Educational Technology* 32 (1) (2001) 5-22
6. Felder R.M. and Silverman L.K.: Learning and teaching styles in engineering education. *Engineering education* 78 (1988) 674-681
7. Gilbert, J.E. and Han, C.Y.: Adapting instruction in search of 'a significant difference. *Journal of Network and Computer Applications* 22 (1999) 149-160
8. Honey, P., Mumford, A.: *The manual of Learning Styles*. Peter Honey Maidenhead (1992)
9. Lu, J., Yu, C.-S., Liu, C.: Learning style, learning patterns, and learning performance in a WebCT-based MIS course. *Information & Management* 0 (2003) 497-507
10. McLoughlin, C.: The implications of the research literature on learning styles for the design of instructional material. *Australian Journal of Educational Technology* 15 (3) (1999) 222-241
11. Papanikolaou, K.A., Grigoriadou, M., Kornilakis, H., Magoulas, G. D.: Personalizing the inter-action in a Web-based educational hypermedia system: the case of INSPIRE. *User-Modeling and User-Adapted Interaction* 13(3) (2003) 213-267
12. Pask, G.: Styles and strategies of learning. *British Journal of Educational Psychology* 46 (1976) 128-48
13. Riding, R. and Rayner, S.: *Cognitive Styles and Learning Strategies*. David Fulton Publishers London 1998
14. Sadler-Smith, E.: 'Learning Style': frameworks and instruments. *Educational Psychology* 17 (1 and 2) (1997) 51-63
15. Solomon, B.S. *Inventory of learning styles*. North Carolina State University (1992)
16. Souto, M.A.M., Verdin, R., Wainer, R., Madeira, M., Warpechowski, M., Beschoren, K., Zanella, R., Correa, J.S., Vicari, R.M., Palazzo de Oliveira, J.: Towards an Adaptive Web Training Environment Based on Cognitive Style of Learning: an Empirical Approach. In: DeBra, P., Brusilovsky, P., Conejo, R. (eds.): AH 2002, LNCS 2347 (2002) 338-447
17. Specht, M. and Opperman, R.: ACE-Adaptive Courseware Environment. *The New Review of Hypermedia and Multimedia* 4 (1998) 141-161.
18. Stern, M.K. and Woolf, B.P.: Adaptive Content in an Online Lecture System. In: Brusilovsky, P., Stock, O., Strapparava, C.: *Adaptive Hypermedia and Adaptive Web-based Systems*, LNCS 1892 (2000). 227-238
19. Stoyanov, S., Aroyo, L., Kommers, P.: Intelligent Agent Instructional Design Tools for a Hypermedia Design Course. In: Lajoie, S. P., Vivet, M. (eds.): *Artificial Intelligence in Education*. IOS Press (1999) 101- 108
20. Triantafyllou, E., Pomportsis, A., Demetriadis, S.: The design and the formative evaluation of an adaptive educational system based on cognitive styles. *Computers & Education* 41 (2003) 87-103
21. VanLehn, K.: Student Modeling. In: Polson, M. C., Richardson, J.J. (eds.): *Foundations of Intelligent Tutoring Systems*. Lawrence Erlbaum Associates Hillsdale NJ (1988) 55-78
22. Vermunt, J.D.H.M.: Metacognitive, cognitive, and affective aspects of learning styles and strategies: A phenomenographic analysis. *Higher Education* 31 (1996) 25-50
23. Witkin H., Moore C., Gooddenough D, Cox P.: Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research* 47 (1977) 1-64

Table II. Comparison of AEHS that use learning style as a source of adaptation

System	Domain	Learning Style Model	Adaptation based on Learning Style	Diagnosis Approach & Dynamic Adaptation
ACE (Specht & Opperman, 1998)	Domain Independent	Learning style preferences: <i>preferences about sequencing of learning materials</i>	Sequencing of learning materials according to a particular teaching strategy (learning by example, reading texts or learning by doing), based on learner's interests and material preferences.	Dynamic adaptation of the teaching strategy is based on info coming from monitoring learners' requests on ed. material, and on the success of the currently used strategy (determined by learner's performance in tests).
CS383 (Carver et al., 1999)	Computer Systems	Sensing/intuitive, visual / verbal, and sequential/global (Felder and Silverman, 1988)	Lesson media elements presented in a sorted list ranked from the most to least conducive based on learners' learning style.	During the first lesson learners submit the questionnaire proposed by (Solomon, 1992) to identify their learning style.
Arthur (Gilbert and Han, 1999)	Computer Science Programming	Learning style preferences: <i>style of instruction during which learners exhibit satisfactory performance</i>	Alternative styles of instruction differ in the type of media they utilize: visual-interactive, auditory-text, auditory-lecture, and text style.	Dynamically adapts the instructional style according to learner's performance in the tests s/he submits.
AES-CS (Triantafillou et al., 2003)	Multimedia Technology Systems	Field dependent / Field Independent (Witkin et al., 1997)	Adapt amount of control (program vs. learner control), contextual organizers (advance vs. post), instructional support, navigational tools and feedback to assessment questions.	Learners submit the Group Embedded Figures Test (GEFT) questionnaire. Direct manipulation of LM (Learner Model).
INSPIRE (Papanikolaou, et al., 2003)	Computer Architecture	Activists, Pragmatists, Reflectors, Theorists (Honey & Mumford, 1992)	Adapt the method and order of presentation of multiple types of educational resources within educational material pages.	Learners submit the questionnaire proposed by (Honey & Mumford, 1992) or they define their learning style. Direct manipulation of LM.
MANIC (Stern & Woolf, 2000)	Domain Independent	Learning style preferences: <i>media (graphic, text), type of instruction, level of content abstractness, ordering of different types of content</i>	Presentation of content objects using <i>stretchtext</i> which allows certain parts of a page to be opened or closed. Sequencing of content objects for a concept based on learner's preferences.	The system dynamically adapts the content presentation by observing learner's interactions with the system.

End-User Quality of Experience Layer for Adaptive Hypermedia Systems

Cristina Hava Muntean and Jennifer McManis

Performance Engineering Laboratory, Faculty of Engineering and Computing,
Dublin City University, Glasnevin, Dublin 9, Ireland,
{hava, mcmanis}@eeng.dcu.ie

Abstract. In the context of new devices and the variety of network technologies that allow access to the Internet, the deployers of Web applications need to ensure that end-users have a positive experience using new applications and they will be willing to re-use them. User experience is dependent not only on the content served to them, but also on the performance of that service. This paper explores a new dimension of individual differences between Web users: end-user Quality of Experience (QoE). It proposes a solution on how to provide satisfactory end-user QoE in the field of educational adaptive hypermedia. A new QoE layer for Adaptive Hypermedia is introduced that attempts to take into account multiple factors affecting Quality of Experience, which might arise from a wide range of Web components (e.g. text, images, video, audio). Usability evaluation based on comparison of a classic adaptive e-learning system with a QoE-aware one has shown that students considered the QoE-aware system significantly more usable than the classic system. Learning performance tests indicated that the changes made by the QoE-aware system did not affect the learning capabilities offered by the classic system.

1 Introduction

Extensive research in the area of Web-based adaptive hypermedia has demonstrated the benefit of providing personalized content and navigation support for specific users or users categories. Web users differ in skills, aptitudes and preferences for processing the accessed information, and goals. They may have different perceptions of the same content and performance factors. Finally, they may have special needs due to disabilities. Therefore, many Web adaptive hypermedia applications have been proposed that try to capture and analyse these user related features in order to optimise the user experience with the Web site.

Many of these adaptive hypermedia systems have been applied in the educational area. This research area has attracted huge interest due to its capability for facilitating personalized e-learning, its distributed nature and its simplicity of interaction.

With the advance in computer and communications technology a variety of Internet access devices (e.g. laptop, pocketPC, PDA, mobile phone) have been launched on the market. The type and capacity of the access device, the bandwidth and the state of the network the device operates on, the complexity of the Web pages delivered over a given network all affect the quality of experience for the end-user. Thus, end-users of educational and training services expect not only high-quality and efficient educational material but also a per-

fect integration of this material with the day-to-day operational environment and network framework. In this context it is significant to highlight a new problem faced by network-based education over the Internet: providing a good level of end-user perceived Quality of Service (QoS), also called Quality of Experience (QoE).

Currently Adaptive Hypermedia Systems for Education (AHSE) place very little emphasis on QoE and its affect on the learning process. This QoE-unaware approach is perhaps unsuited to a general learning environment where one can imagine a student with a laptop moving from a low bandwidth home connection, to a higher bandwidth school connection, and potentially to public transport with a mobile connection with a widely varying bandwidth connection. It should be noted that some AHSs have taken into consideration some performance features, such as device capability [1], in order to improve the end-user perceived quality. However, these account for only a limited range of factors affecting performance (e.g. serving pages in smaller segments if the user has a palmtop rather than a PC) and do not fully address QoE.

Therefore, adaptive hypermedia systems should also take into consideration QoE characteristics when the user profile is built and regularly monitor in real-time any change in the system that might indicate variations of QoE. These include changes in the user's operational environment and also modifications of user behaviour, which might possibly indicate dissatisfaction with service (such as an abort action). This would allow for better Web content adaptation that suites varying conditions.

This paper presents an approach that introduces a new QoE layer to the classic adaptive hypermedia system architecture that would improve the end-user perceived QoS by taking into consideration different performance factors that may affect the end-user satisfaction. This layer provides a Performance Monitor which measures a variety of performance metrics in order to learn about the Web user's operational environment characteristics, about changes in network connection between the user's computer and Web server, and the consequences of these changes on the user's quality of experience. This information is synthesized in a Perceived Performance Model, which proposes strategies for tailoring Web content in order to improve QoE.

To demonstrate the benefits of the proposed QoE enhancement we have deployed it in the open-source AHA! system [2] creating QoSAHA. In this paper we present the results of preliminary tests comparing both performance and usability of AHA! and QoSAHA when delivering an adaptive tutorial in a low bit rate environment. These results indicate that QoSAHA improves performance and user satisfaction with their experience while not affecting the user learning outcome.

2 Quality of Experience (QoE) for Web Applications

The term Quality of Experience focuses on the user and tries to understand end-user expectations for QoS. QoE is considered in [3] as the collection of all the perception elements of the network and performance relative to expectations of the users. The QoE concept applies to any kind of network interaction such as Web navigation, multimedia streaming, voice over IP, etc. According to the type of application the user interacts with, different QoE metrics that assess the user's experience with the system in term of responsiveness and availability have been proposed. QoE metrics may have a subjective element to them and

may be influenced by any sub-system between the service provider and the end-user. ITU-T Recommendation G.1010 [4] provides guidance on the key factors that influence Quality of Service (QoS) from the perspective of the end-user (i.e. QoE) for a range of applications that involves voice, video, images and text. The Recommendation provides a list of parameters that govern end-user satisfaction for these applications.

In the area of World Wide Web applications, QoE has been referred as end-to-end QoS or end-user perceived QoS. Measuring end-to-end service performance, as it is perceived by end-users is a challenging task. Previous research [5, 6, 7] shows that many QoS parameters such as end-to-end response time or page download time, perceived speed of download, successful download completion probability, user's tolerance for delay, and frequency of aborted connections factor into user perception of World Wide Web QoS. Measurement of these parameters may be used to assess the level of user satisfaction with performance. The interpretation of these values is complex—varying from user to user and also according to the context of the user task.

For example, according to a number of studies, user's expectation on the download time is influenced by different contextual factors (e.g. the type of task performed by the user, the duration of time the user interacts with the site, the user's awareness of the connection capabilities) [5, 8]. Currently, there are no standard thresholds for the download time. However, on average a download time higher than 10-12 seconds causes disruption and users lose their attention to the task, while values higher than 30 seconds cause frustration. At the same time it is significant to mention that when the user is aware of his slow connection, he/she is willing to tolerate a threshold of 15 seconds.

End-user perceived QoS has also been addressed in the area of multimedia streaming. Research such as [9, 10, 11] assesses the effect of different network-centric parameters (i.e. loss, jitter, delay), the continuous aspect of multimedia components that require synchronization, or the effect of multimedia clip properties (i.e. frame size, encoding rate) on end-user perceived quality when streaming different type content.

In this paper QoE is addressed only in the area of AHS with applicability in education. Typical e-learning systems may involve a combination of text, images, audio and video, and their quality of service is based on the combination of all of these rather than any individual component. The educational context also has its own set of requirements and user expectations and it is against these that user perceptual quality should be evaluated.

3 QoE Layer Enhancement for AHS

Starting from a generic architecture of an AHS that consists of a domain model (DM), a user model (UM), an adaptation model (AM), and an AHS engine [12] we have enhanced the system with an end-user perceived QoS layer [13, 14]. This QoE layer includes the following new components (see Figure 1): the Perceived Performance Model (PPM), the Performance Monitor (PM), and the Perceived Performance Database (PP DB). The PM monitors different performance metrics that may affect the QoE (e.g. end-to-end response time, round-trip time, throughput, and even user behaviour such as abort actions) in real time during user access sessions and delivers them to the PPM. The PPM models this information and suggests Web content characteristics (e.g. the number of embedded objects in the Web page, dimension of the base-Web page without components and the total di-

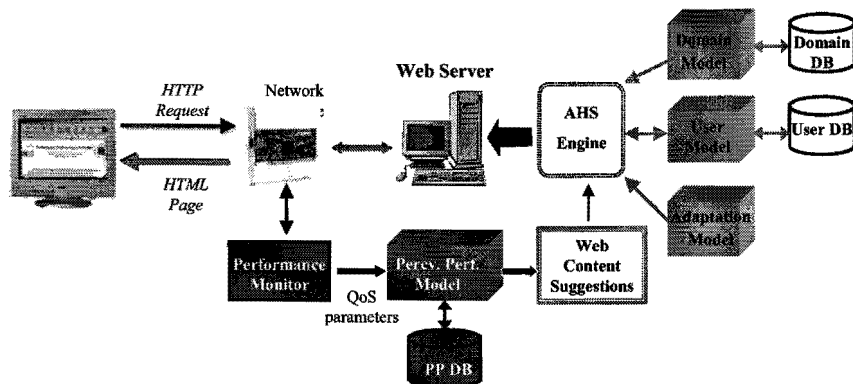


Fig. 1. QoE-aware AHS Architecture

mension of the embedded components) that would best meet the end-user expectation related to QoS. These constraints are applied to the Web pages that have already been designed according to the user profile (based on the UM and AM). The PPM model can also take into consideration the users subjective opinion about their QoE explicitly through the use of a form, which asks users to rate their current QoE. This introduces a degree of subjective assessment specific to each user. A more detailed description of the Perceived Performance Model is presented in [13]. PP DB saves user related performance information.

Applying the PPM suggestions involves the alteration of the properties of the embedded images (that are presented as concepts in the DM) or the elimination of some concepts expressed through text, images, paragraphs or other Web page items. These actions would be applied to those concepts the user is least interested in as recorded by the UM.

For Web pages that consist of text and images the alteration /elimination of images would bring the biggest improvement for the access time. This is due to the fact that images represent the largest percentage of the total size of a Web page. For the situation when audio and video components are part of a Web page, strategies that involve size and quality adjustments for audio and video can be applied (e.g. for video compression techniques involving frame rate, resolution and colour depth modifications and respectively for audio silence detection and removal technique). These techniques are studied by the multimedia networking area and they are not detailed in this paper.

3.1 A Simple Example

This section presents a simple example of applying PPM suggestions in the case of a Web page being downloaded in a low bit rate environment. As PM indicates that the download time is too long (e.g. greater than 10 seconds), the PPM will seek to reduce the amount of data sent to a calculated value that determine an acceptable download time.

Step 1: Image Compression. The first step involves the use of an image compression technique that would reduce the size of the images. Different degrees of compression (expressed as percentage of the original) are applied on each image depending on the user knowledge or interest in the concept represented by the image. If one of the computed compression rates cannot be applied to an image due to the fact that the image compression technique has reached the compression threshold that ensures good quality, an image elimination strategy must be applied.

Step 2: Image Elimination. It is based on image removal from the Web page and its replacement with a link to the image. Consequently, if a user does really want to see the image, the link offers this possibility. The algorithm is based on the following rules:

- the image with the lowest interest for the user is replaced with a link
- if the new recomputed total size of embedded objects from a Web page is still higher than the PPM suggestion, perform again step 1 (image compression). In this case a lower compression rates will be applied in the remained images.

4 Assessing the Benefits of the QoE Layer Using QoSAHA

For illustration and testing purposes the end-user perceived QoS related enhancements have been deployed on the open-source AHA! system, creating QoSAHA. The AHA! system was developed at the Eindhoven University of Technology, in the Database and Hypermedia group. It was first deployed and used in educational area as an adaptive hypermedia courseware application that supports the "Hypermedia Structures and Systems" course [15, 16]. The following advantages of AHA! allowed us to use it and to demonstrate their benefits by performing different tests using the AHA! tutorial:

- The AHA! has been extensively tested and accepted by the research community.
- The AHA! system architecture respects the generic AHS architecture [17].
- AHA! is open source.
- AHA! version 2.0 includes an adaptive tutorial as example of the adaptive features of the AHA! system.

4.1 Evaluation of the QoSAHA

Although many AHSE have been proposed and developed, there is a significant lack of evaluation strategies and comprehensive empirical studies to measure the usefulness and effectiveness of adaptation within the systems and between the systems. There is also much debate on how adaptive hypermedia applications should be evaluated since there is no standard or agreed evaluation framework for measuring the value and the effectiveness of adaptation yielded by adaptive systems. In order to determine the evaluation strategies for QoSAHA system an extensive survey of the research in the adaptive education area with emphasis on Web-base AHSE has been undertaken [18].

We compare the proposed QoSAHA with the original AHA! system both in terms of performance and end-user perception. Simulation tests are used to determine access times for both systems in different bit rate environments. This provides the basis of our performance comparison. Subjective tests are used to compare end-user perceived QoS, user satisfaction, and user learning capabilities for the two systems. These tests have assessed the following evaluation criteria: time taken to complete a task, learner achievement and performance, and usability.

4.2 Simulation Tests

Reducing the access time of the Web pages involved in a learning process can produce a significant improvement into the end-user QoE. The simulation tests involve comparative measurements of the access times for the AHA! and QoSAHA systems respectively when a learning task is performed by the user in different operational environments (home modem connection, broadband connection - ISDN and LAN connection) and with various connection throughputs (from 28 kbps to 128 kbps and over). The learning task involved the study of the "AHA! installation" section from the AHA! tutorial consisting of four Web pages. In order to comply with the Web content constraints generated by the PPM, only image compression techniques needed to be used in these tests.

The simulation tests show that QoSAHA system improves the total access time of the learning process for users in low bit rate environments (64 kbps and lower) by up to 37% with a reduction in the quantity of data sent of up to 42%. A subsequent user survey suggests this reduction did not significantly affect the quality of the images. An average of up to 9.7% perceived quality degradation was reported for the lowest bit rate (28 kbps) but still close to the "good" perceptual level. The results are detailed in [14].

4.3 Subjective Evaluation

The goal of the subjective evaluation is to compare the learning outcome for users using QoSAHA and AHA! systems and to assess user satisfaction with the two systems. The evaluation tested both learner achievement and usability. The conditions used were the same as those used for the simulation tests (which showed an increase in measured performance using QoSAHA). For our preliminary tests we showed that QoSAHA also improved user QoE without affecting the learning outcome.

4.3.1 Setup Conditions

The laboratory-network setup used for testing involved four desktops PC Fujitsu Siemens with Pentium III (800MHz) processors and 128 MB memory, a Web server IBM NetFinity 6600 with two processors Pentium III (800 MHz) and 1GB memory and one router Fujitsu Siemens with Pentium III (800MHz) processor and 512 MB RAM that has a NISTNET network emulator installed on it. The NISTNET instance that allows for the emulation of various network conditions characterized by certain bandwidth, delay and loss rate and pattern was used to create a low bit rate operational environment with a 56 kbps modem connection (Figure 2). This setup offers similar connectivity to that experienced by residential users and is the same as that used in the simulation tests.

The subjects involved in this study are comprised of forty-two postgraduate students from Faculty of Engineering and Computing at Dublin City University. They were asked to complete a learning task that involved the study of the "AHA! installation" chapter from the AHA! tutorial. The subjects were randomly divided into two groups. One group used the original AHA! system, whereas the second used QoSAHA. No time limitation was imposed on the execution of the learning task.

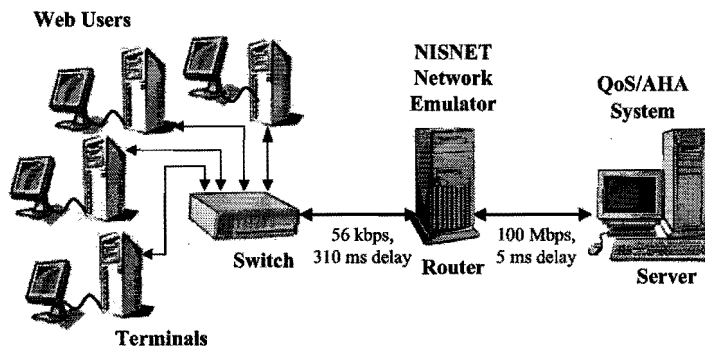


Fig. 2. Laboratory-network configuration for the perceptual testing

At the start of the study session the subjects were asked to read a short explanation concerning the use of the system and the required duties. Their duties were as follows:

- *complete a Pre-Test* that consists of a questionnaire with six questions related to the learning topic. The test is used to determine subject's prior knowledge in this domain
- *log onto the system* and proceed to browse and study the material
- *complete a Post-Test* at the end of the study period. The Post-Test consists of a questionnaire with fifteen questions that test recollection of facts, terms and concepts from the supplied material, as suggested in Bloom's taxonomy [19].
- *answer a Usability questionnaire* that consists of ten questions categorized into navigation, accessibility, presentation, perceived performance and subjective feedback.

In order to fully assess the subjects learning achievement, both pre-test and post-test questionnaires were devised from the four different types of test-items most commonly used in the educational area: Yes-no, Forced-choice, Multi-choice and Gap-filling test items. These test items have different degrees of difficulty and their corresponding answers have been assigned weights in the final score accordingly. The maximum score for pre-test is 10 points and the maximum score for post-test is 30 points. The final scores were normalized in the range of 0 to 10.

4.3.2 Learner Achievement

Learner achievement is defined as the degree of knowledge accumulation by a person after studying a certain material. It continues to be a widely used barometer for determining the utility and value of distance learning technologies.

Table 1. Pre-Test results

	mean score	min score	max score	sdev
AHA!	0.35	0	2	0.55
QoSAHA	0.3	0	2	0.53

Table 2. Post-Test results

	mean score	min score	max score	sdev
AHA!	6.70	4.3	9.3	1.401
QoSAHA	7.05	4.6	9.0	1.395

We have performed an analysis of the learner achievement in terms of the final scores from the Pre-test and Post-test achieved by the subjects using the QoSAHA and AHA! systems. The results of the Pre-Test and Post-Test are shown in Table 1 and Table 2.

A t-test two-sample analysis, with equal variance assumed, applied on the Pre-Test scores shows that statistically both groups had the same prior knowledge of the studied

subject (significance level $\alpha=0.01$, $t=0.21$, $t_{critical}= 2.42$, $p(t)=0.41$). According to the Post-Test results the mean score of the subjects that used QoSAHA was 7.0 and that for those that used AHA! 6.6. T-test analysis suggests that this difference in mean does not indicate a significant difference in performance between the two groups of users. ($\alpha=0.05$, $t=-0.79$, $t_{critical}=1.68$, $P(t)=0.21$).

Since the answers for three questions from the post-test questionnaire required the subjects study the images embedded in the studied pages, an analysis of the students' performance on these questions was performed. After the scores related to these three questions were normalized in the range 0 to 10, the mean value of the students' scores was 6.3 for the QoSAHA group and 6.4 for AHA! group. A t-test two-sample analysis, with equal variance assumed again indicates that there is no significant difference in the students performance ($t=-0.08$, $t_{critical}=2.71$, $p(t)=0.93$, confidence level $\alpha=0.01$) although a slight degradation in the image quality was applied by the QoSAHA system.

In summary, preliminary test results indicate the QoE aware AHA! system did not affect the learning outcome, offering similar learning capabilities as classic AHA! system.

4.3.3 Usability Assessment

At the end of the study session both group of subjects were asked to complete an online usability evaluation questionnaire consisting of ten questions with answers on a five-point scale (1-poor - 5-excellent). The questions were created using the widely used guidelines suggested by Preece [20] for evaluating the Web sites. The questions were categorized into navigation presentation, subjective feedback, accessibility, user perceived performance. The last two question categories seek to assess the end-user QoE. Four questions of our survey relate to these two categories. These four questions assess user opinion in relation to the overall delivery speed of the system (Q6), the download time of the accessed information in the context of his/her experience with Web browsing (Q7), the user's satisfaction in relation to the perceived QoS (Q9) and whether the slow access to the content has inhibited them or not (Q5). The presentation of the results on the QoE related questions for the AHA! and QoSAHA systems is shown in Figure 3.

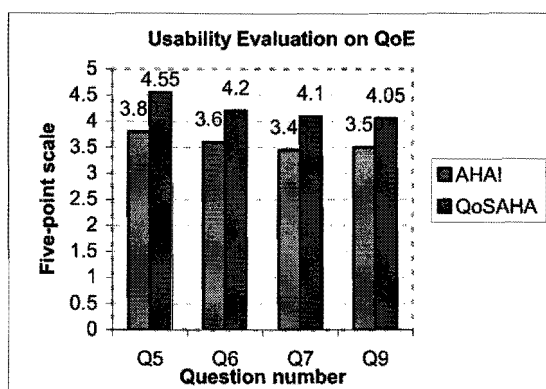


Fig. 3. Usability evaluation results on questions that assessed the end-user QoE

As seen from the chart the QoSAHA system has provided a better QoE for the end users, improving their satisfaction to a “good” level in relation to using the system. This was even though the subjects were using a slow connection (56 kbps) during the study and they were not explicitly informed about this. A t-test two-sample analysis on the results of these four questions confirmed that users’ opinion on QoE is significantly higher for QoSAHA than for AHA!, stated with confidence level above 99%, ($p < 0.01$).

The usability assessment on the other questions related to the navigation and presentation features achieved an average score of 3.83 for AHA! and 3.89 for QoSAHA, demonstrating that these features were not affected by the addition of the QoE enhancements.

An overall view of the results of usability testing of both systems, when all ten questions were considered of equal importance shows that the students considered QoSAHA system (mean value=4.01) significantly more usable than AHA! one (mean value=3.73). These results were also confirmed by the unpaired two-tailed t-test ($t=2.44$, $p < 0.03$) with a 97% confidence degree. This achieved 7.5% increase in the overall QoSAHA usability mainly due to higher scores obtained in the questions related to end-user QoE.

5 Conclusions

This paper has proposed Quality of Experience (QoE) as another dimension of user characterisation that should be taken into consideration by the personalization process provided by adaptive hypermedia applications. QoE is directly influenced by the operational environment through which the user interacts with the AHS (bandwidth, delay, loss, device capabilities, etc) and by the subjective assessment of the user of perceived performance. The goal of any AHS should be not only to provide the content that would best suit the user’s goals, knowledge, or interest but also to provide the best content that would fit the user’s operational environment. In this context we have proposed a QoE-layer enhancement for AHS that analyses some key factors that influence QoE and makes a correlation between their values and Web page characteristics that provide the best QoE for the end-user.

For evaluation purpose QoSAHA was created by deploying the QoE layer on the open-source AHA! system used in the educational area. This paper presents a study on the impact that the incorporation of the QoE layer has on learning outcome and on usability of the system in a low bit operational environment (56 kbps). This study is based on a performance comparison between the AHA! system and QoSAHA and involves both simulations and subjective testing.

The evaluation results show that the learning outcome was not affected by the deployment of the QoE enhancement. Most significantly they also suggest that with the price of a slight degradation in quality of the images, an important improvement in system usability was provided by QoSAHA system. This was mainly due to improvements in user satisfaction related to their QoE.

The next stage of the QoE layer evaluation is to perform learning performance and usability assessment for the case of image elimination technique when increase compression would produce high degradation into the image quality. Other evaluation criteria such as time taken to complete a learning task or time taken to search for a term, how many times

the subjects have revisited some Web pages during the learning session will be assessed as part of further QoSAHA evaluation.

Acknowledgements

Informatics Research Initiative of Enterprise Ireland support is gratefully acknowledged.

References

1. Smyth B., Cotter P.: Content Personalisation for WAP-enabled Devices. Workshop Machine Learning in the New Information Age: Conference on Machine Learning, Barcelona (2000)
2. AHA! Project <http://aha.win.tue.nl/>
3. Assuring QoE on Next Generation Networks, White paper, Empririx
4. ITU-T Recommendation G.1010, End-user multimedia QoS categories, November 2001
5. Bhatti, N., Bouch, A., Kuchinsky A.: Integrating user-perceived quality into Web server design. *Computer Networks Journal* Vol. 33, No. 1-6, (2000), 1-16
6. Krishnamurthy, B., Wills, C.W.: Analysing factors that influence end-to-end Web performance. *Computer Networks Journal*, Vol. 33, No.1-6, (2000), 17-32
7. Bouch, A., Kuchinsky, A., Bhatti, N.: Quality is in the eye of the beholder: Meeting users' requirements for Internet Quality of Service. In *Proceedings of the ACM CHI 2000 Conference on Human factors in computing systems*, Hague, Netherlands (2000)
8. Ramsay, J., Barbesi, A., Preece, J: Psychological investigation of long retrieval times on the World wide Web. *Interacting with Computers Journal*, Vol. 10 No. 1 (1998)
9. Blakowski G., Steinmetz, R.: A Media Synchronisation Survey: Reference Model, Specification, and Case Studies, *IEEE Journal on Selected Areas in Communications*, vol. 14 (1) 1996, 5-35
10. Ghinea G. and Thomas, J. P: QoS Impact on User Perception and Understanding of multimedia Video Clips, in *Proc. of ACM Multimedia '98*, Bristol, UK, (1998), 49-54.
11. Watson A., Sasse, M. A.: Multimedia conferencing via multicasting: determining the quality of service required by the end user, in *Proc. of AVSPN '97*, Aberdeen, UK, (1997), 189-194.
12. Wu H., De Kort E., De Bra, P.: Design issues for general-purpose adaptive hypermedia systems. *12th ACM Conf. on Hypertext and Hypermedia*, Aarhus, Denmark (2001) 141-150
13. Muntean, C., H., McManis, J.: A QoS-aware adaptive Web-based system, *IEEE International Conference on Communications (ICC04)*, Paris, France (2004)
14. Muntean, C., H., McManis, J.: QoSAHA: A performance oriented learning system, *AACE ED-MEDIA Conference*, Lugano, Switzerland (2004)
15. TU/e course 2L690 Hypermedia Structures and Systems, <http://wwwis.win.tue.nl/2L690/>
16. De Bra, P.: Teaching Through Adaptive Hypertext on the WWW. *Journal of Educational Telecommunications*, Vol. 3, No. 2/3 (1997) 163-180
17. De Bra, P., Calvi, L. : AHA! An open Adaptive Hypermedia Architecture. *The New Review of Hypermedia and Multimedia*, Vol. 4, (1998) 115-139
18. Muntean, C. H., McManis, J.: Adaptive e-learning systems: evaluation issues. *Transactions on Automatic Control and Computer Science*, Vol. 49 (63), 2004, ISSN:1224-600X
19. Bloom, B. S., Mesia, B. B. and Krathwohl, D. R: *Taxonomy of Educational Objectives* (two vols: The Affective Domain & The Cognitive Domain). New York. David McKay Inc (1964)
20. Preece, J.: *Online Communities: Designing Usability, Supporting Sociability*. In John Wiley & Sons (eds), Chichester UK (2000)

Evaluating Presentation Strategy and Choice in an Adaptive Multiple Intelligence Based Tutoring System

Declan Kelly¹, Brendan Tangney²

¹ National College of Ireland, Dublin, Ireland
dkelly@ncirl.ie

² University of Dublin, Trinity College, Ireland
tangney@tcd.ie

Abstract. EDUCE is an Intelligent Tutoring System for which a set of learning resources has been developed using the principles of Multiple Intelligences. It can dynamically identify learning characteristics and adaptively provide a customised learning material tailored to the learner. This paper describes a research study using EDUCE that examines the relationship between the adaptive presentation strategy, the level of choice available and the learning performance of science school students aged 12 to 14. The paper presents some preliminary results from a group of 18 students that have participated in the study so far. Results suggest that learning strategies that encourage the student to use as many resources as possible are the most effective. They suggest that learning gain can improve by presenting students initially with learning resources that are not usually used and subsequently providing a range of resources from which students may choose.

1 Introduction

Research on learning shows that students learn differently, that they process and represent knowledge in different ways, that it is possible to diagnose learning style and that some students learn more effectively when taught with appropriate strategies [24]. EDUCE [15][16] is an Intelligent Tutoring System for which a set of learning resources has been developed using the principles of Multiple Intelligences [10]. It can dynamically identify user learning characteristics and adaptively provide customised learning material tailored to the learner [17]. The multiple intelligence concept defines intelligence as the capacity to solve problems or fashion products that are of value and states that there are different ways to demonstrate this intelligence. It is a concept that offers a framework and a language for developing a broad range of content that supports creative, multi-modal teaching. In EDUCE four different intelligences are used to develop four categories of content: verbal/linguistic, visual/spatial, logical/mathematical and musical/rhythmic intelligences. Currently, science is the subject area for which content has been developed.

This paper describes an empirical study that examines the relationship between the adaptive presentation strategy, levels of choice and the learning performance of

science school students aged 12 to 14 in a computer based adaptive learning environment using different versions of EDUCE. The adaptive presentation strategy involves guiding students to resources they prefer and do not prefer to use. The level of choice is determined by the range of resources a student has access to and the extent by which the student is guided to a particular resources by EDUCE.. Learning performance is defined by learning gain, learning activity and motivation. Learning gain is measured by a pre and post test, learning activity is determined by the navigation profile and motivation by the attempts to answer questions.

The goal of the research study is to address the following research questions:

- Does providing a range of learning resources improve learning gain and activity?
- What are the advantages in making adaptive presentation decisions in relation to giving the learner complete control?
- What is the difference between learning gain and activity when presenting resources that are preferred and resources that not preferred?

The results of this study may be significant for researchers and practitioners. For researchers, it will produce specific results that demonstrate the relationship between learning and the availability of different learning resources. For practitioners, it demonstrates how teaching in different ways can affect learning.

2 Background

Research has provided a wealth of insight into individual differences and orientations to learning that can be translated into instructional design [7],[11]. Different studies report that when the learner's individual learning style is taken into account, the quality of the learning is enhanced [8], [20], [23]. However, observing and defining learning characteristics is difficult and traditionally questionnaires and psychometric tests are used to assess and diagnose learning characteristics [14], [22].

Gardner's Multiple Intelligence (MI) [11, 12] concept is a psychological theory that addresses what the brain does with information. It defines intelligence as the capacity to solve problems or fashion products that are of value. It states that there are eight different ways to demonstrate this intelligence with each having its own unique characteristics, tools, and processes that represent a different way of thinking, solving problems, and learning. Its use in the classroom has been significant [3] but its application to online learning and intelligent tutoring systems is still undergoing research. However, even though it is a theory and has no specific application method or instructional approach it does offer a structure and language in which to inform the student, domain and pedagogical model of an intelligent tutoring system.

Several systems adapting to the individual's learning characteristics have been developed [4], [5]. However in developing such systems it is not clear which aspects of learning characteristics are worth modelling, how the modelling can take place and what can be done differently for users with different learning styles [2]. In attempts to build a model of student's learning characteristics, feedback from the

student is obtained using questionnaires, navigation paths, answers to questions, directly requesting feedback, allowing the user to update their own student model and to make specific adaptations such as sorting links or viewing stretch text.

Machine learning techniques offer one solution in the quest to build a model of learning characteristics [26], [27]. Typically these systems contain a variety of instructional types such as explanations or example and fragments of different media types representing the same content, with the tutoring system choosing the most suitable for the learner. Another approach is to compare the students performance in tests to that of other students, and to match students with instructors who can work successfully with that type of student [13]. Other systems try to model learning characteristics such as logical, arithmetic and diagrammatic ability [19].

EDUCE builds on existing research by modelling the learner's psychological characteristics through the multiple intelligence concept.

3 Student Model

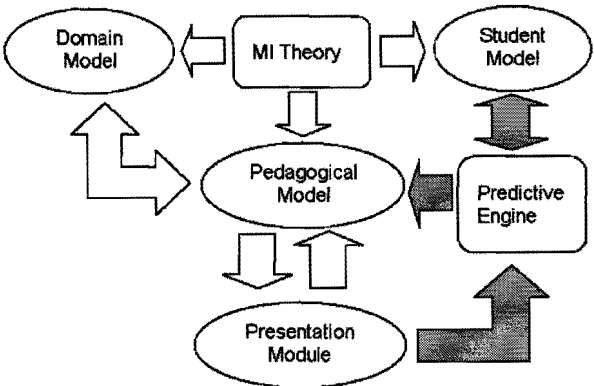


Figure 1. Educe Architecture

Fig. 1 illustrates the architecture of EDUCE. It consists of a student model , a domain model, a pedagogical model, a predictive engine and a presentation model. The predictive engine receives input from the presentation module, builds the student model and informs the pedagogical model.

The MI concept inspires the student model in EDUCE. Gardner identifies eight intelligences involved in solving problems, in producing material such as compositions, music or poetry and other educational activities. The intelligences include the logical/mathematical, linguistic/verbal, visual/spatial, bodily/kinesthetic, musical/rhythmic, interpersonal, intrapersonal and naturalist. Currently EDUCE uses the following four intelligences in modeling the student and in the future will be extend to incorporate the other intelligences. The four intelligences is use are

- Logical/Mathematical intelligence (LM) - This consists of the ability to detect patterns, reason deductively and think logically.
- Verbal/Linguistic intelligence (VL) - This involves having a mastery of the language and includes the ability to manipulate language to express oneself.
- Visual/Spatial intelligence (VS) - This is the ability to manipulate and create mental images in order to solve problems.
- Musical/Rhythmic intelligence (MR) - This encompasses the capability to recognise and compose musical pitches, tones and rhythms.

EDUCE builds a model of the student's learning characteristics by observing, analysing and recording the student's choice of MI differentiated material. Other information also stored in the student model includes the navigation history, the time spent on each learning unit, answers to interactive questions and feedback given by the student on navigation choices.

In particular the model that describes how a student uses different resources is built using the following criteria:

- Did the student spend a minimum amount of time using the resources ?
- Did the student spend a long time using the resource ?
- Which resource did the student use first ?
- Did a student use only one resource or multiple resources ?
- Did the student use the resource more than once ?
- Did the student attempt a question after viewing the resource ?
- Did the student attempt a question after viewing the resource and get it right ?

4 Domain Model

The domain model is structured in two hierarchical levels of abstraction, concepts and learning units. Concepts in the knowledge base are divided into sections and sub-sections. Each section consists of learning units that explain a particular concept. Each learning unit is composed of a number of panels that correspond to key instructional events. Learning units contain different media types such as text, image, audio and animation. Within each unit, there are multiple resources available to the student for use. These resources have been developed using the principles of Multiple Intelligences. Each resource uses dominantly one intelligence and is used to explain or introduce a concept in a different way. To access each of the intelligences, there is a set of practical techniques, methods, tools, media and instructional strategies.

MI is a theory with a set of principles. It structures and suggests but does not prescribe a particular pedagogical model or set of instructional strategies. Moving from a theory of intelligence to actual implementation is an act of interpretation and there has been a considerable amount of research done in articulating different techniques that can access each of the intelligences [3, 18]. Figure 2 shows EDUCE's model for developing MI material [16]. It describes the range of instructional approaches that will cultivate each of the intelligences.

For example, to emphasis the logical and mathematical mind, strategies described as number, order, logic, representation, puzzles, problem solving , relationships, compare/contrast and outlining may be employed.

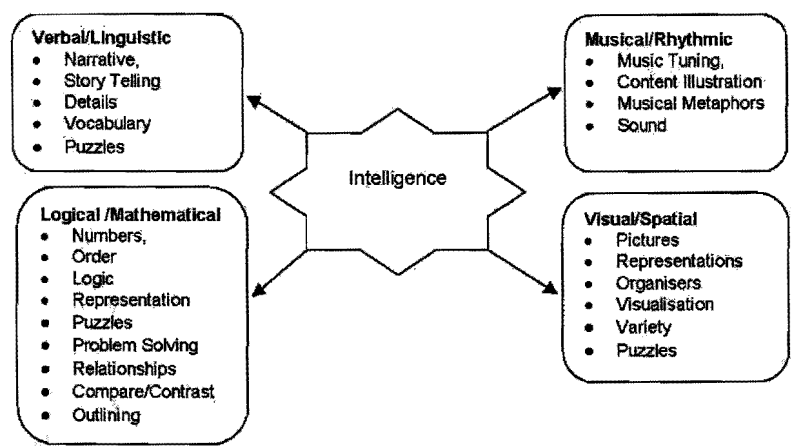


Figure 2. Pedagogical Model for developing MI material

Number includes the use of mental arithmetic, calculations and measurements that encourages mental maths, numerical thinking and precision. The arrangement and detection of order can be promoted through the identification of steps, procedures, sequences and patterns. Logic includes the use of scientific, deductive and inductive logic. This can be best realised by examining how reasoning processes operate and how truthful conclusions may be reached. Syllogisms, venn diagrams and analogies may be employed. Visual representation through the use graphs, charts, piecharts, tables, grids, matrices can make mathematical relationships easier to understand. Mathematical representation involves the use of abstract symbols, codes and formulas to represent and communicate concrete objects and concepts. Logic puzzles and games can awaken and arouse reasoning and logical thinking. Problem solving may be promoted through the use of estimation, prediction, exploration and heuristics. Understanding causal knowledge involves the use of questioning, creating meaningful connections between ideas and understanding cause and effect. Classification and the arrangement of information into rational frameworks includes comparing and contrasting concepts, attribute identification, categorisations and ranking. Outlining explains concepts in logical frameworks using logical explanations, logical thought maps and sequence charts.

Figure 3 gives a specific example of how these instructional strategies were used in the tutorial Static Electricity. Content has been developed in the subject area of Science for the age group 12 to 14.

All resources developed have been assessed and validated by expert MI practitioners.

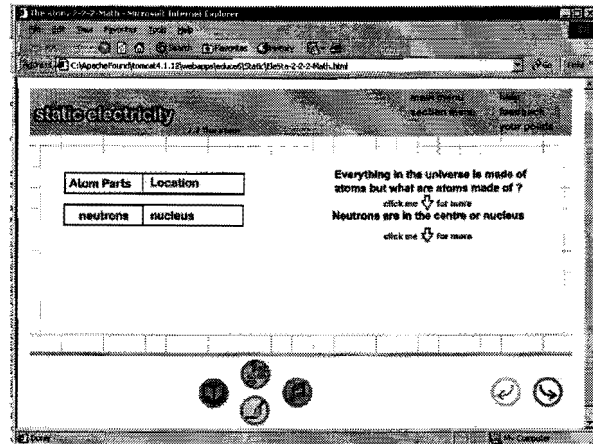


Figure 3. Logical/Mathematical Intelligence

5 Presentation Module

In the teaching of a concept, key instructional events are the elements of the teaching process in which learners acquire and transfer new information and skills. The EDUCE presentation model has four key instructional events, as shown in Fig. 2. [9]

- *Awaken*: The main purpose of this stage is to attract the learner's attention.
- *Explain*: Different resources reflecting MI principles are used to explain or introduce the concept in different ways.
- *Reinforce*: This stage reinforces the key message in the lesson
- *Transfer*: Here learners convert memories into actions by answering interactive questions

At the Awaken stage, to progress onto the next panel, the learner chooses one from four different options. Each choice will lead to a different resource that predominately reflects the principles of one intelligence. At the Reinforce and Transfer stage the learner has the option of going back to view alternative resources.

Different learners may prefer a different sequence events during the learning process, but the for the purpose of determining which resources a learner prefers the instructional event model is the same for all. The navigation path is designed to force the student to make a conscious choice about which resource is preferred. As learners choose between different MI resources, EDUCE automatically builds a model of the learning characteristics and strengths.

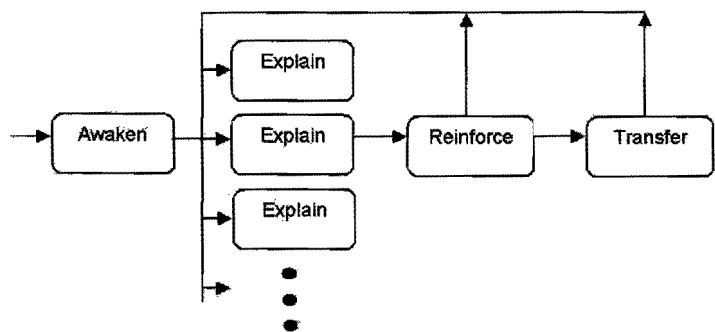


Figure 2. EDUCE Instructional Event Model

5 Pedagogical Model

The pedagogical model uses the predictive engine to make predictions about which MI differentiated resource a student prefers. Being able to predict student behaviour provides the mechanism by which instruction can be adapted and by which to motivate a student with appropriate material. As the student progresses through a tutorial, each learning unit offers four different types of resources. The prediction task is to identify at the start of each learning unit which resource the student would prefer. These predictions are used in two ways: to guide the student to preferred resources and to guide the student to resources that are not preferred. The predictive engine is implemented using the Naïve Bayes algorithm [4]. In EDUCE the current focus of the pedagogical model is on content presentation and selection. Later versions will accommodate concept sequencing.

6 Research Design

Different configurations of EDUCE support different instructional strategies and levels of choice. The research purpose of the experiment described in this paper is to explore the relationship between the independent variables: instructional strategy and level of choice, and the dependent variable: learning performance. Other variables such as MI profile, gender, previous computer experience and level of ability in school will also be examined. The experiment is intended to provide insight into the advantages and disadvantages of providing a range of resources and of guiding learners to preferred resources.

The instructional strategy or more specifically the presentation strategy for delivery material encompasses two main strategies.

1. *Most preferred*: - showing resources the student prefers to use
2. *Least preferred*: - showing resources the student least prefers to use

For each learning unit, there are four MI based learning resources. Which of these resources is shown first is determined by the dynamic and static MI profile and the instructional strategy. The static MI profile of each student is determined before the experiment using an MI inventory (MIDAS)[25]. EDUCE also builds a dynamic model of the student's MI profile by observing, analysing and recording the student's choice of MI differentiated material. Other information also stored in the student model includes the navigation history, the time spent on each learning unit, answers to interactive questions and feedback given by the student on navigation choices.

The second independent variable is the level of choice. There are four different levels of choice provided to different groups. These include:

1. *Free* – student has the choice to view any resource in any order. No adaptive presentation decisions are made as the learner has complete control.
2. *Single* – student is only able to view one resource. This is determined by EDUCE based on an analysis of the MI inventory completed by the student.
3. *Multi* - student is first given one resource but has the option to go back and view alternative resources. The resource first given to the student is determined by EDUCE based on the analysis of the MI inventory completed by the student. The *Multi* choice level is the same as the *Single* choice level but with the option of going back and viewing alternative resources.
4. *Adaptive* – the student is first given one resource but has the option to go back and view alternative resources. The resource first given to the student is determined adaptively by EDUCE. The predictive engine within EDUCE [17] identifies the most preferred and least preferred resource from the online student computer interaction.

Learning performance is defined mainly by learning gain, learning activity and motivation. To calculate the learning gain each student before and after a tutorial will sit a pre-test and post test. The test for the pre-test and post-test is the same and consists of questions that appear during the tutorial. Learning activity is determined by the navigation profile. It is a measure of the different panels visited, the number of different resources used, the reuse of particular resources and the direction of navigation. Learning motivation or persistence is a measure of the student's progression within the tutorial and the attempts made to answer questions. The questions are multi-choice question with four options. Both learning activity and motivation are analysed to provide informed explanations on learning gain.

Students have been randomly assigned to one of the four groups defined by the levels of choice. Each student sits through two tutorials. They will experience both instructional strategies of least preferred and most preferred. To ensure order effects are balanced out, students are randomly assigned to systematically varying sequence of conditions. The design of the experiment can be described as a mixed between/within subject design with counterbalance.

7 Preliminary Results and Future Work

One group of students consisting of 18 boys with an average age of 13 has participated in the study so far. The group was divided into two sets, A and B. Set A use the Free choice version of EDUCE and had the choice to use resources in any order. Set B used the Adaptive choice version of EDUCE where the presentation decisions were made based on the dynamic student profile. Notwithstanding the number of participants, the results were analysed to determine differences between the free choice version and the adaptive choice version. For the adaptive version, the results were analysed to determine any differences between the least preferred and most preferred strategy.

A one-way between groups analysis of variance on the learning gain both on the first day and on the second day was conducted to explore the impact of adaptivity and non-adaptivity. The three groups consisted of one group using the free choice version and two groups using the adaptive least/most preferred versions. There was no significant difference between the groups at the $p < 0.05$ level. On inspecting the mean increase in learning gain for the different versions it interestingly reveals that the mean for the adaptive least preferred version was greater than the free choice version which was in turn greater than that of the adaptive most preferred version on both days.

To evaluate the impact of learning strategy on learning gain within subject, a paired-samples t-test was conducted. There was a statistically significant increase in learning gain using the least preferred strategy ($M=26.25$, $SD=13.024$) and in using the most preferred strategy ($M=10$, $SD=11.95$), $t(7)=2.489$, $p < 0.042$). The eta squared statistic (0.47) indicates a large effect size. This initial surprising result suggests that learning gain increases where students do not get their preferred learning resource. However on closer examination of the learning activity, it is found that students when given their least preferred learning resource increase their learning activity and are exposed to a wider range of resources. It suggests that strategies that increase learning activity and develop all faculties are effective in increasing learning gain.

Currently, the empirical study is underway and on completion more than 200 students will have participated in the research study.

References

1. Armstrong, T. (1993): *Multiple Intelligences in the Classroom*, Penguin Group.
2. Brusilovsky, P. (2001): *Adaptive Hypermedia. User Modeling and User-Adapted Instruction*, Volume 11, Nos 1-2. Kluwer Academic Publishers.
3. Campbell, L. & Campbell, B. (2000): *Multiple Intelligences and student achievement: Success stories from six schools*, Association for Supervision and Curriculum Development.
4. Carver, C., Howard, R., & Lavelle, E. (1996): *Enhancing student learning by incorporating learning styles into adaptive hypermedia*. 1996 ED-MEDIA Conference on Educational Multimedia and Hypermedia. Boston, MA.

5. Danielson, R. (1997): Learning Styles, media preferences, and adaptive education. In: Proceedings of the workshop "Adaptive Systems and User Modeling on the World Wide Web", Sixth International Conference on User Modeling, UM97.
6. Duda, R. & Hart, P. (1973): Pattern Classification and Scene Analysis, Wiley, New York.
7. Dunn & Dunn (1978): Teaching Students through their individual learning styles: A practical approach. Prentice Hall, Reston Publishing.
8. Ford, N & and Chen, S. Y (2001): Matching/mismatching revisited: an empirical study of learning and teaching styles. British Journal of Educational Technology, Vol 32 No 1.
9. Gagné, R. M., Briggs, L. & Wagner, W., (1992). Principles of instructional design. Harcourt, Brace.
10. Gardner H. (1983) Frames of Mind: The theory of multiple intelligences. New York. Basic Books
11. Gardner H. (1993): Multiple Intelligences: The theory in practice. New York. Basic Books.
12. Gardner H. (2000): Intelligence Reframed: Multiple Intelligences for the 21st Century. Basic Books
13. Gilbert, J. E. & Han, C. Y. (1999): Arthur: Adapting Instruction to Accommodate Learning Style. In: Proceedings of WebNet'99, World Conference of the WWW and Internet, Honolulu, HI.
14. Honey, P. and Mumford, A. (1986, 1992): The Manual of Learning Styles. Peter Honey, Maidenhead.
15. Kelly, D. & Tangney, B. (2002): Incorporating Learning Characteristics into an Intelligent Tutor. In: Proceedings of the Sixth International on ITSs, ITS2002.
16. Kelly, D. & Tangney, B. (2003): A Framework for using Multiple Intelligences in an ITS. In: Proceedings of EDMedia'03.
17. Kelly, D. & Tangney, B. (2004). Predicting Learning Characteristics in a Multiple Intelligence based Tutoring System. In: Proceedings of the Seventh International on ITSs, ITS2004..
18. Lazaer, D. (1999): Eight Ways of Teaching: The Artistry of Teaching with Multiple Intelligences, SkyLight.
19. Milne, S. (1997): Adapting to Learner Attributes, experiments using an adaptive tutoring system. Educational Psychology Vol 17 Nos 1 and 2, 1997
20. Rasmussen, K. L. (1998): Hypermedia and learning styles: Can performance be influenced? Journal of Multimedia and Hypermedia, 7(4).
21. Reigeluth, C.M. (1996): A new paradigm of ISD ? Educational Technology, 36(3)
22. Riding, R. J. (1991): Cognitive Styles Analysis, Learning and Training Technology, Birmingham.
23. Riding, R., & Grimley, M. (1999): Cognitive style and learning from multimedia materials in 11-year children. British Journal of Educational Technology, 30(1).
24. Riding, R. & Rayner, S. (1997): Cognitive Styles and learning strategies. David Fulton Publishers.
25. Shearer, C. B. (1996): The MIDAS handbook of multiple intelligences in the classroom. Columbus. Ohio: Greyden Press.
26. Specht, M. and Oppermann, R. (1998): ACE: Adaptive CourseWare Environment, New Review of HyperMedia & MultiMedia 4,
27. Stern, M & Woolf, B. (2000): Adaptive Content in an Online lecture system. In: Proceedings of the First Adaptive Hypermedia Conference, AH2000.

Considering Human Memory Aspects to Adapting in Educational Hypermedia*

Peter Ágh and Mária Bieliková

Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
bielik@fiit.stuba.sk
<http://www.fiit.stuba.sk/~bielik>

Abstract. In this paper we target the limited capacity of the human memory while developing adaptive educational hypermedia systems. We discuss implications of remembering and forgetting for the adaptive hypermedia systems development. The forgetting is characterized as a consequence of time passed between two learning events. Knowledge from psychology is used for stating implications of the human memory properties for an improvement of the adaptive learning systems. An experimental implementation of the model of remembering and forgetting is described.

1 Introduction

Current adaptive educational hypermedia (AH) systems recognize several aspects of an individual user such as user's goals/tasks, knowledge, background, preferences, interests, or user's individual traits [4]. Important aspect considered in educational AH systems is undoubtedly a level of the user's knowledge related to the learned topic (in the IEEE Personal and Private Information [8] learner profile denoted as the learning performance). The user model reflects current state of the user knowledge related to the presented information as it is comprehended by the AH system. The user's characteristics change (evolve) in the course of learning in accordance with changes of current state of his knowledge (as evaluated by the AH system).

Most current AH systems assume that the amount of user knowledge only grows. But increasing knowledge (as a consequence of the remembering) is not the only process. The user can also *lose* (e.g., forget) some knowledge. The remembered knowledge is not stored in the human memory forever but in the course of time the knowledge can (and some of them will) drop out from the memory.

* This work has been supported by the Grant Agency of Slovak Republic grant No. VG1/ 0162/03 "Collaborative accessing, analysis and presentation of documents in Internet environment using modern software tools".

Considering mentioned characteristics of the learning process is important during the learning [10]. We presume that a utilization of the human memory aspects while developing an educational AH system would also improve the effectiveness of the AH system usage through an improvement of the learning process. Assume for example the following situation: the adaptive book "presumes" that a user possesses adequate knowledge (prerequisites) for understanding a concept just explained. In spite of truly learned concept some time ago, now – after some time passed from this learning event – the user forgot some of the previously acquired knowledge (because of long time without any repeating). The knowledge forgetting causes inconsistencies between the user model as represented in the AH system (which does not consider the remembering and forgetting in an adaptation of the educational material to the individual) and the actual state of the user's knowledge. As a result, we will likely observe incorrect recommendation of the educational AH system.

Described situation occurs due to not considering specific characteristics of the human memory. In this paper we describe some issues related to the human memory and implications for adaptive hypermedia. We consider the human memory as a new aspect of the user's background modelled in the AH system user model. We give several suggestions for increasing effectiveness of the AH system, especially educational AH systems. In the paper we presume some "minimal amount of knowledge" delivered to the user via the AH system because the effect of the knowledge forgetting process becomes significant with only relatively large knowledge spaces.

The rest of the paper is organized as follows. In the Section 2, we briefly present known facts from psychology about the human memory and the processes of remembering and forgetting. In the next section, we discuss implications of the human memory characteristics for adaptive hypermedia and propose a model, which considers the human memory characteristics. Finally, conclusions and further directions of our research are stated in the Conclusions.

2 Background of human memory models

The human mind can be viewed as an information processing system. Its architecture is thought to consist of three basic components: sensory memory, working memory and long-term memory [2]. These components roughly correspond to the input (the human mind perceives information from the outside through the senses), processing (information from the sensory memory is processed in the working memory) and storage (processed information is stored in the long-term memory) (see Fig. 1). Naturally, information stored in the long-term memory can be accessed, or activated to help with the processing in the working memory. Accessing information is perceived as the remembering that can be viewed as a usage of the system (to be able to find information later again). This view provides a useful basis for considering the human memory characteristics during the learning process [10].

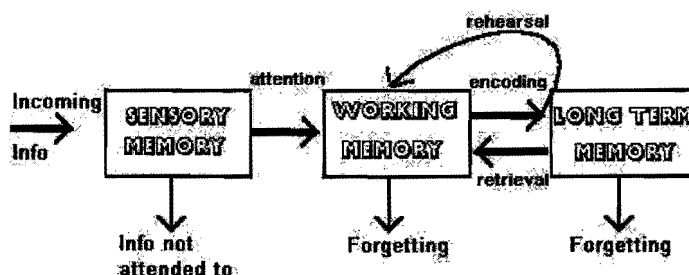


Fig. 1. Model of human information processing system [2].

Information stored in the working memory can be looked up much faster than in the long-term memory. The working memory is essential for reading comprehension. Frequently, the read sentence is related to the previous sentences, so the new sentences are considered along previous according their sense. It is believed that this process is accomplished in the working memory (as a consequence people with higher capacity of working memory are able faster understand a text).

One of the most interesting and significant characteristics related to the human mind is the very small capacity of the working memory known as the magical number "seven plus or minus two" [9]. The limited storage capacity of the working memory is accompanied also by a relatively brief duration (estimates range from 12 to 30 seconds without a rehearsal), which results in the information loss.

The forgetting is viewed primarily as a consequence of

- fading (trace decay) over time,
- interference (overlying new information over the old) or
- lack of retrieval cues.

The information loss can be prevented by means of repeating. Here the elaborative rehearsal which in contrast to maintenance rehearsal involves deep semantic processing of a to-be-remembered information item¹ is more effective [6]. The maintenance rehearsal involves only simple rote repetition aiming at lengthening periods of time the information item is maintained in the working memory. The elaborative rehearsal can be supported by guidelines.

Accessing an information item can be influenced by several factors. Time of searching the information item can be cut down with a good guideline. But the effect of a guideline degrades with the rising number of information items bound with the guideline. Expectant reason is that the system of guidelines brings a hierarchical organization of the information items. The benefit is that the search is performed on the smaller file of information items. However, every new related information item enlarges the file and aggravates the hierarchy.

¹ For example, if an individual is presented with a list of digits for later recall (4968214), grouping the digits together to form a phone number transforms the stimuli from a meaningless string of digits to something that has a meaning.

Mentioned factors imply the forgetting. Function of the volume of remembered information depends on time and has a character of falling an exponential curve. So called the *forgetting curve* was first described by Ebbinghaus in 1885 [7]. To test the retention, Ebbinghaus practiced a list of information items until he was able to repeat the items correctly two times in a row. He then waited varying lengths of time before testing himself again. The forgetting turned out to occur most rapidly soon after the end of practice, but the rate of forgetting slowed as time went on and fewer items could be recalled. It was showed that an individual lost most of the learned information items in first hours (after 8 hours was on average remembered less than 40% information items). After this time is the oblivion less intensive (in average still more than 30% after 24 hours and a bit lesser than 30% after month). Ebbinghaus also discovered that distributing learning trials over time is more effective in memorizing than massing practice into a single session; and he noted that continuing to practice material after the learning criterion has been reached enhances retention.

For the integrity, let us notice that the information items loss can also have biological reasons. It is possible that some biological processes necessary for encoding, storing or searching are disrupted. For example, in a process of embedding knowledge in the memory some structures of brain including hippocamp and amygdala are active. Their mutilation has a negative influence on the process of remembering.

3 Some implications for adaptive hypermedia

While designing an educational adaptive hypermedia system we can take an advantage of knowing characteristics of both working and long-time memories. Considering the human memory along other aspects of the student's background brings several assets:

- more accurate information about the state of the user's (student's) knowledge,
- better aid for remembering the knowledge and automatic repetition of lost knowledge,
- more effective (adaptive) assistance for students with memory-problems, or in opposite, assistance for students with more-than-average memory abilities.

First two items are related to the effectiveness of delivery of the educational material by the AH system in general. The last item enables to accommodate individual differences in the human memory capacity for personalization.

To achieve mentioned assets several techniques can be used. The most important are:

- hierarchical organization of the learned material – improves the information access and enables the effective usage of limited capacity of the working memory,
- guidelines for effective information searching – helps to overcome an interference between the information items,

- forming linkages between the concepts (natural or artificial) – results in a possibility of elaborative rehearsal and better structuring of the knowledge space,
- measuring understanding the sense of knowledge (or measure the usage of the knowledge) – helps in planning of the repeating,
- considering a context of the environment – helps to decrease the amount of information items considered at the time by giving contextual cues.

Some of the techniques are already used in designing educational adaptive hypermedia systems or authoring their content [5].

Following discussion is based on our experiments with the adaptive book on computer architecture where we applied knowledge on the human memory characteristics [1]. Our proposals can be easily incorporated to various adaptive hypermedia educational systems.

3.1 Modeling the process of remembering and forgetting

The simplest method of modeling the remembering and forgetting information items in the student's memory is an application of the forgetting curve [7]. Using only the forgetting curve directly is insufficient because we can only infer how much per cent from the original grist of the information items has been remembered in some point of time. However, we cannot recognize whether specific information item in a given point of time *is remembered* or has been *lost*.

We can say only that the information item (learned at time t and not repeated) is remembered with high probability (if according to data about the user's memory-losing at time t is remembered more than $K\%$ of learned information items, e.g. more than 90%) or lost with high probability (likewise).

We have proposed a simple model which reflects the forgetting. It extends every concept's traditional performance value from the user model [8] with data about how much is the knowledge remembered. We call it *knowledge activity in the memory*. Every knowledge (represented by a concept in the domain model) has defined the knowledge activity in the memory represented by a real number. It's value must be upon given bound B , otherwise the knowledge represented in the concept is considered as being *lost* (from the user's memory). After a successful learning the corresponding concept is set as "learned" and the knowledge activity in the memory is set to a value greater than B . Moreover, after every new user's session with the AH system, the user model reflects the forgetting curve by decreasing the knowledge activity for every concept not being used in the session.

Described approach ensures that the repeated knowledge or knowledge more used are being lost more slowly. The knowledge-remembering model can be supplemented by including hierarchical binds between the knowledge items in a domain.

3.2 Remembering and repeating

Knowledge is remembered better if we work more with it. It is not enough if the information item representing a knowledge only appears many times on a page presented to the student. A measure of remembering depends on how much is the knowledge substantial (e.g., whether it is a prerequisite for understanding another knowledge presented on the page, or whether it is necessary for finding-out results of exercises) and on the appearance (layout) of the corresponding information item.

It is advisable to distinguish at least three levels of the rate at which a user has worked with particular knowledge represented by the information item:

- *normal level*: the user has worked with the knowledge in such a way that after the end of the session he has remembered it and can correctly reproduce it;
- *low level*: the user has worked with the knowledge less than in the normal level (e.g., the information item has been mentioned just a few times among many other information items) and
- *high level*: the user has worked with the knowledge more than in the normal level (e.g., the user has intensively and repetitively worked with the information item and successfully passed several exercises related to the knowledge).

While in the first case the *speed of losing* can be computed according to the standard forgetting curve, in the second case the oblivion is faster and in the last case slower. Of course, there is no linear relationship: very high measure of the user's work with a knowledge does not substantially increase its measure of the remembering. The measure of remembering of a knowledge item depends on a "measure of working" with it. However, the raising is very slow from a certain level. The reason lays in the memory limits. It is possible to remember more than at the normal level (e.g., frequently used knowledge, important knowledge) but not substantially more. On the other hand, if the measure of working has been low (e.g., the information item has been put down only once) the probability of remembering the knowledge is very low.

The same is true not only for *learned knowledge*, but also for the *repetition*. After a knowledge has been learned, its activity in the memory in time decreases. By repeating and using the knowledge, its activity in the memory increases. For example, if a user studies a page where the knowledge item K is referenced or repeated, or should be used for understanding other assertions, all these activities increase the knowledge activity in the memory. Of course, there is also important measure of the user's work. For example, if a knowledge was noticed only (in a text, comment, footnote) or announced, then increasing the knowledge activity in the memory is futile.

The open issue is the determination of a list of the knowledge items considered during the inference related to remembering and repeating. It seems that it is not possible particularize the list automatically. We can count up automatically the frequency of textual representation of a given knowledge in the given text but this does not reflect its "importance". It may happen that the knowledge

has no textual occurrence in the presented page, however, for understanding the content of the page the user should work with the knowledge intensively (so it is repeated many times). On the other hand, another knowledge item may occur frequently in the text but it is not much important for the comprehension (and therefore not repeated so much). Information on the mapping of knowledge items to the content should be provided by the author of the educational material.

3.3 Repetition

Information items the user read on a page are inserted into his working memory. Because of limited capacity of the working memory the information items are either moved to the long-term memory or they are lost. To support the process of moving the information items into the long-term memory (i.e., to enforce the remembering) it is effective to repeat them.

One possibility is a *periodical repetition*. After the user has learned given "amount" of the knowledge, the AH system provides the repetition of the knowledge learned from the previous repetition. The repetition can take several faces. In our adaptive book it is automatically observed how many new knowledge items the user has learned. Providing the summation of the occupied items is greater than the predefined capacity limit the AH system invokes a repetition. The system generates a page with the resume of learned knowledge (occurrence of the knowledge items in the information fragments is tagged by the author). The complexity of the knowledge item is also considered. Described approach does not give exact results, but it ensures a repetition in time closed to the point where the user has learned certain amount of the knowledge.

Other techniques of the repetition realized in our adaptive book are:

- repeat at the end of a lesson the knowledge learned in the lesson (*final repetition*),
- repeat at the beginning of a new lesson the knowledge learned in the previous lesson (*overall introductory repetition*),
- repeat at the beginning of a new lesson the knowledge (assumed) necessary in this lesson (*necessary introductory repetition*).

The same can be applied to sessions or various parts of the book content.

Often it is not practical or possible to repeat all of the knowledge items marked as forgotten. The AH system should select a set of knowledge items for the repetition. Certain number of the knowledge items is selected and only these knowledge items are repeated at the beginning of a new lesson. If there is large number of the lost knowledge items the adaptive book offers a repetition-lesson, aimed for the repetition only.

Selection can be made on several criterions, for example: random selection, selection based on time of the acquisition a knowledge (priority is given to the knowledge acquired longer time ago), selection based on a measure of remembering, i.e. the activity in the memory characteristic is used (priority is given to the knowledge item with lower activity in the memory), selection based on prerequisite-dependencies (priority is given to the knowledge item which is supposed to be in the need of the user in the next study time).

3.4 Knowledge space organization

Knowledge space is formed by the concepts (with corresponding information fragments). The concepts are connected by relations. The currently most used approaches to structuring the knowledge space are the hierarchical approach and the network approach [5]. The structure of the hyperspace can aid the repeating in such a way that the repeating one knowledge item may cause the need of repeating (in part or in whole) another knowledge item. The same holds for the forgetting.

For example, if a student is able to compute the volume of a cylinder, he must be able to compute the square of a number. In opposite, if he has forgotten how to compute the square of a number, he will not be able to compute the volume of a cylinder. But it is not true that if the student has forgotten to compute the volume of a cylinder, he also has forgotten how to compute the square of a number or that if he remembers how to compute the square of a number he also remembers (and knows) how to compute the volume of a cylinder.

The prerequisite relation is well known relation in adaptive educational hypermedia [3]. Considering the human memory characteristics it is useful to distinguish between *domain prerequisites* and *pedagogical prerequisites*. Let A be a prerequisite of B. If A is a domain prerequisite, the student is constrained in understanding B with requirement to understand A. If A is a pedagogical prerequisite the constraint is weaker and it is possible to comprehend B without knowing A. As an example, let us present expressions in C programming language course. If the adaptive book explains this part using commonalities and differences between C and Pascal languages, then the knowledge about Pascal are denoted as pedagogical prerequisites. The student needs a knowledge of Pascal to understand this part of the content. But, when a repetition process is evoked on the "expressions in C" knowledge item, the Pascal knowledge item is not necessary to be repeated.

There can be an objection that the above is not fully true. We may repeat some topics of Pascal when we repeat C language. For example, some things may be the same (or similar) and the user may have remembered data like "in C it is the same like in Pascal". The user may also remember the page itself, text or/and its graphical layout on the page. It is also possible that when he would hear about some topics of C language, he will bring back some information about Pascal. In all these cases the user will repeat with some knowledge about C language also some knowledge about Pascal. This may happen. But after some amount of time the intensity of repeating related knowledge items will decrease and the user will repeat only already repeated knowledge and its domain-depended prerequisites.

4 Conclusions

The research discussed in this paper addresses the possibility of improving effectiveness of learning using adaptive educational hypermedia by considering the human memory characteristics. Important aspect is limited capacity of the

working memory. We discussed impacts of the human mind nature to the adaptive hypermedia systems. Our research is supported by experimental adaptive web-based book. Known adaptation techniques (annotations of links and conditional inclusion of fragments) are supplemented by an inference based on a model of the remembering and the forgetting which leads to the repeating. The base for modelling the remembering is the forgetting curve. The forgetting curve can be tuned individually for each user which results in more effective repeating by utilization of individual differences.

We still work on experimental evaluation of issues elaborated in this paper. Our future work will concentrate on using experiments for proving effectiveness of the proposed approach. Naturally, we expect that the proposed models should be tuned for particular usage and differences of the individuals.

References

1. Peter Ágh. Efficient learning using adaptive hypermedia, December 2001. Diploma Thesis supervised by Mária Bielíková. Slovak University of Technology.
2. Richard Atkinson and Richard M. Shiffrin. Human memory: A proposed system and its control processes. In K. Spence and J. Spence, editors, *The psychology of learning and motivation*. Academic Press, 1968.
3. Paul De Bra, Geert-Jan Houben, and Hongjing Wu. AHAM: A dexter-based reference model for adaptive hypermedia. In K. Tochtermann, J. Westbomke, U.K. Wiil, and J. Leggett, editors, *Proc. of ACM Conference on Hypertext and Hypermedia*, pages 147–156, Darmstadt, Germany, February 1999.
4. Peter Brusilovsky. Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2):87–100, 2001.
5. Peter Brusilovsky. Developing adaptive educational hypermedia systems: From design models to authoring tools. In Stephen Blessing Tom Murray and Shaaron Ainsworth, editors, *Authoring Tools for Advanced Technology Learning Environments*. Kluwer, 2003.
6. F.I.M. Craik and R.S. Lockhart. Levels of processing. a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, (11):671–684, 1972.
7. Hermann Ebbinghaus. *Memory. A Contribution to Experimental Psychology*. Reprinted Bristol: Thoemmes Press, 1999, 1885.
8. IEEE. IEEE P1484.2/D7, 2000-11-28. Draft Standard for Learning Technology. Public and Private Information (PAPI) for Learners. Available at <http://ltsc.ieee.org/>, 2000.
9. George A. Miller. The magical number, seven, plus or minus two: sme limits on our capacity for processing information. *Psychological Review*, (63):81–97, 1956.
10. Juhani E. Tuovinen. Optimising student cognitive load in computer education. In Ainslie E. Ellis, editor, *Proc. of the Australasian conference on Computing education*, pages 235–241. ACM Press, 2000.

Discovery of Individual User Navigation Styles

Eelco Herder and Ion Juvina

Eelco Herder

Department of Computer Science, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands
herder@cs.utwente.nl

Ion Juvina

Institute of Information and Computing Sciences, Utrecht University
Padualaan 14, De Uithof, 3584 CH Utrecht
ion@cs.uu.nl

Abstract. Individual differences have been shown to lead to different navigation styles. In this paper we present a pilot study that aims at finding predictors for users' vulnerability to experience disorientation that can be gathered unobtrusively and in real-time. We identified two navigation styles that we called flimsy navigation and laborious navigation that together predict users' perceived disorientation. Our findings suggest that adaptive navigation support that addresses these navigation styles is a promising means to ease the various problems that are commonly associated with users experiencing disorientation.

1 Introduction

Individual differences, ranging from gender differences through system experience to cognitive styles, significantly influence the way that people navigate through hypermedia systems [5]. Many of these individual user characteristics can be gathered using questionnaires or standardized tests. However, for adaptive hypermedia systems this approach is often undesirable, as it requires time and effort from the users, which might eventually put them off. Moreover, not all user characteristics are stable or easily measurable: as an example, a user's motivation and concentration is most likely to change over time.

For this reason, it makes sense to provide users with adaptive navigation support based on users' navigation styles [8]. With knowledge of the strategies that users follow, it is easier to recognize patterns in their navigation paths that indicate usability problems that need to be solved. A typical usability problem is that users become disoriented, or *lost* in a web site [18], which means that they are unable to keep track of their positions: at some point users might not know where they are, how they came there or where they can go to. Several characteristics of user navigation, most importantly those related to page revisits, have been related to success measures, such as task outcomes and user's perceived disorientation [5][8][11].

In this paper we present the results of a pilot study that was aimed at finding patterns in user navigation that indicate a user's vulnerability to perceive

disorientation while working on goal-directed tasks that require a fair amount of navigation to complete them. We were able to extract two navigation styles – which we called *flimsy navigation* and *laborious navigation* - that performed well in predicting the user's perceived disorientation. In the next section we will describe shortly how individual differences influence user navigation. Navigation styles and measures for user navigation are dealt with in the subsequent section. The presentation of the pilot study and its results will be followed with a discussion on the generalizability of the study and the implications for adaptive navigation support.

2 Individual Differences in Web Navigation

There is a vast amount of literature showing and analyzing individual differences involved in web navigation. In [7] it is noticed that novices tend to make use of a linear structure in hypermedia systems, when it is made available, while experts tend to navigate non-linearly. [10] demonstrated that students who had more *domain knowledge* displayed more purposeful navigation and allocated time more variably to different pages. *Spatial ability* is an important determinant of hypermedia navigation performance, as reported in several studies [e.g. 4]; users with low spatial abilities have difficulty in constructing and using a visual mental model of the information space. Students with an internal *locus of control* are reported to be better able to structure their navigation and to take advantage of hypertext learning environments [10].

Research on cognitive mechanisms involved in web navigation gains increasing influence in the HCI community. A cognitive model of web navigation should be able to simulate the navigation behavior of real users, producing the same navigation patterns as actual users would do. Many approaches to user navigation modeling are mostly inspired by the theory on *information foraging* [13]. Information foraging theory assumes that people, when possible, will modify their strategies in order to maximize their information gain. More specifically, users continuously compare the benefits of alternative actions, for example digging further into one information resource versus looking for a different resource. Process models that are based on these theories can analyze or simulate users' actions in terms of their individual evaluations of their expected utility.

3 User Navigation Styles

A related line of research aims at directly modeling the user's navigation behavior in order to provide adaptive navigation support in web applications [8]. A dynamic user navigation model could include:

- *syntactic information* (e.g. which links are followed, what does the navigation graph look like, what is the time that users spent on each page)
- *semantic information* (i.e. what is the meaning of the information that the user encountered during navigation)

- *pragmatic information* (i.e. what is the user using the information for, what are the user's goals and tasks)

In this section we focus on the syntactic information. Our aim is to identify patterns in user navigation that indicate problems associated with disorientation, as experienced by the user. In the first subsection we characterize several user navigation styles. In the second subsection we introduce several measures that can be used to capture these navigation styles.

3.1 Navigation Styles and Page Revisits

User navigation can range from goal-directed task completion to more unstructured browsing and exploration of the availability of information or services [7]. Routine browsing is an integral part of web navigation, nowadays; typically, users have a small collection of favorite sites that they visit very frequently [5]. Several taxonomies of web browsing behavior are presented in the literature. One of the finer grained taxonomies is presented in [15], a white paper that is clearly targeted at the ecommerce community in which seven patterns are categorized, based on session length, average page view times and the amount of revisits during this session.

Within a navigation session, users often return to pages that serve as *navigational hubs*. Extensive use of these hubs is reported to be an effective navigation strategy [11]. When looking for information, users often employ search strategies that are quite similar to graph searching algorithms, such as *depth first*, *breadth first* and *heuristic search* [2].

With knowledge of the type of session that users are involved in, and the navigation styles that they employ during these sessions, it is possible to recognize navigation patterns that might indicate usability problems.

3.2 Measures of User Navigation

User navigation paths can be modeled as graphs, with the vertices representing the pages visited and the edges representing the links followed [8]. Several – mostly graph-theoretic and statistical – methods can be used for analyzing this structure. Typical measures include the total number of pages visited to solve a task, the total time needed to solve a task and the average times spent on single pages [2]. Within the navigation paths, patterns may exist that indicate a user navigation style or problems encountered. In our pilot study we made use of a collection of navigation measures that together describe these patterns. They will be shortly described below. For a more detailed discussion about these measures we refer to [8].

Number of Pages and Revisits

As mentioned before, page revisits are very common in web navigation. By capturing various aspects of page revisitation, we aim to find revisitation *patterns* rather than the *amount* of revisitation. The following measures were taken into account:

- the **path length** is the number of pages that the user has requested during a navigation session, including page requests that involved revisits;

- the **relative amount of revisits** is calculated as the probability that any URL visited is a repeat of a previous visit. We adopted the formula that is suggested by Tauscher and Greenberg [17];
- the **page return rate** indicates the average number of times that a page will be revisited. The return rate is calculated by averaging the number of visits to all pages that have been visited at least twice. A more extensive use of navigation landmarks will most likely lead to a limited set of pages that is visited very frequently;
- **back button usage** indicates the percentage of back button clicks among the navigation actions, including backtracking multiple pages at once using the back button;
- **relative amount of home page visits** is a self-descriptive label. ‘Relative’ refers to a correction of home page visits based on path length.

View Times

The *average time* that users spend at web pages is reported to be an important indicator for user interest and human factors [16]. Besides the average view time, the *median view time* was also taken into account, as users generally spend only little time on the large majority of pages before selecting a link [3]. The median view time is not affected by the few ‘high content’ pages that were inspected more carefully, and thus provides a better indicator for the average view time while *browsing*.

Navigation Complexity

Navigation complexity can be defined as ‘any form of navigation that is not strictly linear’. Complexity measures are mostly derived from graph theory and used frequently for assessing hypertext and its usage [8]. Typical measures reflect the cyclical structure of the navigation graph and the length of navigation sequences within the graph. Several commonly used complexity measures were taken into account:

- the **number of links followed per page** (‘fan degree’) [14] represents the ratio between the number of links followed and the number of distinct pages visited;
- the **number of cycles** [14] is calculated as the difference between the number of links followed and the number of pages visited. As the number of cycles grows with the length of the navigation path, it can only be used for a fixed time window;
- the **path density** [14] compares the navigation graph to the corresponding fully connected graph. A higher path density indicates that a user makes use of short navigation sequences and regularly returns to pages visited before;
- **compactness** [11] is a measure similar to path density. It indicates that users follow a ‘shallow’ search strategy. In contrast to the path density, it compares the average distance between any two pages in the navigation graphs to a theoretical minimum and maximum;
- the **average connected distance** [3] indicates the average length of a path between any two connected pages in the navigation graph. A higher average connected distance indicates that users do not return to a page very soon, but only after having browsed for a while. They also return using a link rather than using the back

button. In short, the average connected distance measures the users' confidence in that they 'will find their way back later'.

The navigation measures that are described above are labeled *first-order measures* in this paper, because they are derived directly from the raw data, without taking into account that the measures might be correlated, which most likely would be the case. As an example, the *average connected distance* is calculated independently of *back button usage*, without taking into consideration the fact that usually low values on the former measure are associated with high values on the latter and vice-versa. This aspect was dealt with by calculating *second-order measures* – or *navigation styles*, as will be explained in the next section.

4 Pilot Study – Navigation Styles and Disorientation

In our pilot study we were interested in what navigation styles occur when users perceive disorientation when performing several goal-oriented tasks. In order to better interpret the outcomes, we also collected several user characteristics – as introduced in section two – as well as users' evaluation of their navigation activities. The experimental setup and the results will be discussed in this section.

4.1 Experimental Setup

The study consisted of individual sessions with thirty subjects, all undergraduate and graduate students from two Dutch universities in the age range 19-28, with an average age of 21.5. Participants were selected randomly out of the student lists of both universities, while making sure that males and females were equally presented.

Each session consisted of three stages:

- collection of data on user characteristics;
- the actual navigation session and collection of navigation data;
- evaluation of the navigation session, including a survey on users' perceived disorientation.

Several user characteristics were collected in the first stage. The characteristics that are relevant in the context of this paper are briefly described below. For more details we refer to [9]. *Spatial ability*, *episodic memory* and *working memory* were measured with computerized cognitive tests provided by the Dutch research institute TNO Human Factors. The users' *internet expertise* is composed of self-reported frequencies of internet use and self-assessed level of knowledge. At the beginning of the navigation session the users rated their affective disposition; users who rated themselves high on the states determined, calm and alert, and low on the states sluggish and blue, were considered to be in an *active mood*. *Locus of control* refers to the users' belief in how they contributed to their own success or failure, which was measured with a 20-item scale.

In the navigation session, subjects were asked to perform various tasks in the field of web-assisted personal finance. This field includes using the web to keep a personal

budget, to perform financial transactions and decide to save or invest money. The tasks were designed in such a way that it would require a fair amount of navigation to answer or to solve them [9]. Subjects had thirty minutes in total to solve the tasks. Three web sites were used in this study, two of which are dedicated to personal finance. They provide users with advice and tools – such as planners, calculators and educators – to deal with their financial problems. The third site, an online store, was used as a reference.

After the navigation session, the subjects were asked to evaluate their satisfaction with task completion and the usability of the different web sites used. A survey on perceived lostness [1] was also included in the evaluation session.

4.2 Results

As it is most likely that patterns in the first order navigation measures occur simultaneously, second order navigation measures – linear combinations of the first order measures – were calculated. Principal component analysis with *equamax* rotation on twenty-two navigation measures resulted in four factors that together explained 86% of the variance. We will focus on two factors, which account for 27% and 23%, respectively, after rotation. We labeled them *flimsy navigation* and *laborious navigation*, based on their correlations with first-order measures and user characteristics. It should be noted that these styles do not exclude one another. All correlations mentioned are significant with $p < 0.05$.

High scores on the *flimsy navigation style* are associated with:

- small number of pages visited ($r = -0.80$)
- high path density ($r = 0.80$)
- high median view time ($r = 0.77$)

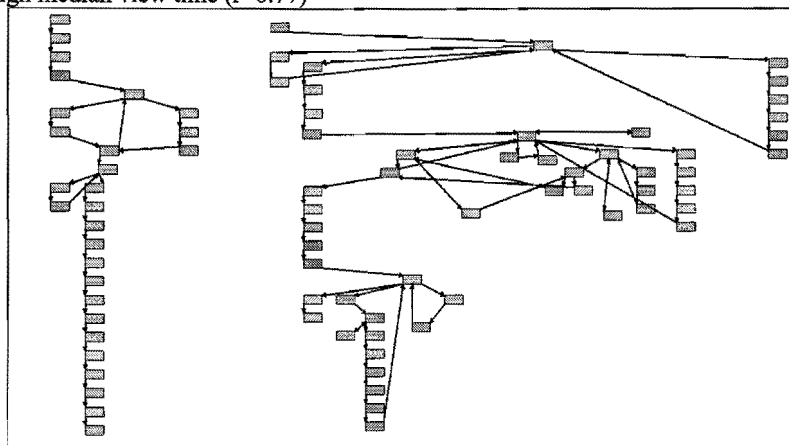


Figure 1. Flimsy (left) versus sturdy (right) navigation. From the figure it can be observed that flimsy navigation is characterized by short navigation paths and a low number of cycles in the navigation graph. The page revisits that did take place in the flimsy navigation path were made using the back button

- short average connected distance ($r=-0.70$)
- low number of cycles ($r=-0.53$)
- high rate of home page visiting ($r=0.48$)
- high frequency of back button use ($r=0.39$).

Flimsy navigation appeared to be a weak navigation style. Most of the navigation takes place around the site's home page and users regularly return to their starting points. Time is mostly spent on processing content instead of actively locating information. The short average connected distance indicates that users return to a page very soon. Users also prefer to return by using the back button instead of by following links. The low number of cycles indicates that users employing this navigation style do not make extensive use of the means for revisitation available within the sites.

High scores on the flimsy navigation style are associated with low scores on Internet expertise, current active mood, working memory and locus of control. Based on these correlations, it is likely that flimsy navigation is mostly employed by inexperienced users who are not able or not inclined to reconstruct their past actions; rather, they continue along the same path or eventually start over again. For these reasons, we might expect that flimsy navigation is related to users' perceived disorientation.

High scores on the *laborious navigation style* are associated with:

- high number of links followed per page ($r=0.95$)
- high revisitation rate ($r=0.94$)
- high number of cycles ($r=0.79$)
- high return rate ($r=0.73$)
- high frequency of back button use ($r=0.71$)
- high path density ($r=0.43$)
- high number of pages visited ($r=0.40$)
- short average connected distance ($r=-0.39$).

This navigation style involves intensive exploration of navigational infrastructure provided by the site. Users seem to employ a trial and error strategy; they follow links merely to see if they are useful or not. They figure out quite fast when paths are not leading towards their goal and return. Revisits are numerous but not redundant: once a page is revisited a different link is followed than before, which constitutes another trial.

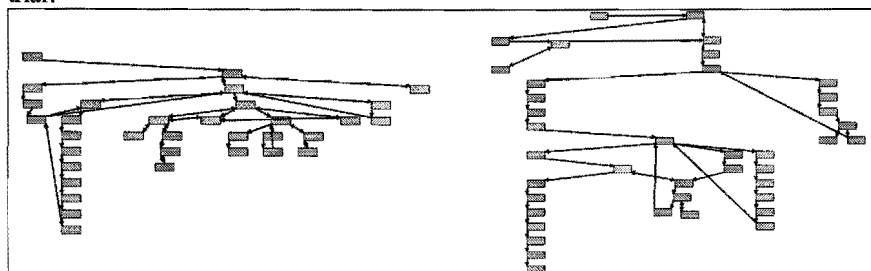


Figure 2. Laborious (left) versus non-laborious (right) navigation. From the figure it can be observed that the laborious navigation style is characterized by a high amount of revisits, with some pages clearly functioning as navigational landmarks.

This behavior is particularly observed on navigational hubs, such as menus and index pages.

High scores on *laborious navigation* are associated with high *episodic memory*, and low *spatial ability*. This style indicates a revisitation pattern that does not lead to disorientation; instead, laborious navigation appears to help users in constructing a conceptual overview of the site structure and then to make use of this model.

Multiple linear regression analysis was used to find out which navigation measures and navigation styles performed best in predicting the subjects' perceived disorientation. Including predictors in regression models was based on the *stepwise method*; the predictive power must be seen as the best one can get with the minimum number of predictors. It turned out that the *flimsy* and *laborious* navigation styles together best predicted the user's perceived disorientation ($R^2=0.29$) with a large effect size ($ES^2 = 0.29/0.71 = 0.41$)¹.

Table 1. Prediction of perceived disorientation based on navigation styles. The regression model consists of *perceived disorientation* as dependent variable and *flimsy navigation* and *laborious navigation* as predictors. From the regression coefficients (B) one can observe the positive and negative correlations of flimsy and laborious navigation respectively with perceived disorientation. The standardized coefficients (Beta) show a larger relative importance of flimsy navigation as compared with laborious navigation.

	B	Beta	t	Sig.
(Constant)	40.1		29.66	0.000
Flimsy navigation	3.92	0.46	2.85	0.008
Laborious navigation	-2.38	-0.28	-1.73	0.095

5 Discussion

The results of our pilot study suggest that users' vulnerability to experience disorientation in large web sites can be automatically diagnosed with an attractive level of accuracy. We identified two navigation styles, *flimsy navigation* and *laborious navigation*, which proved to be significant predictors with a large effect size.

The area in which these navigation styles have been identified, is rather limited: they apply to situations where goal-directed and performance-oriented tasks are performed on the web. The domain of web assisted personal finance might seem narrow and this is why we used three different web sites and a relatively complex and heterogeneous range of task. By choosing three different websites to be used in the pilot study, we attempted at randomizing factors pertaining to a specific site structure or interface design. Tasks were not only aiming at locating information but also at

¹ The effect size for regression is calculated with the following formula: $ES^2 = R^2/(1-R^2)$. 0.02 is considered a small effect, 0.15 a medium one and 0.35 a large effect size [6].

using this information to solve actual problems. These decisions were intended to constitute premises for ecological validity and generalizability of the results.

The number of subjects (thirty) was rather limited and relatively homogenous, as they were students. New data is necessary to find out in what situations the identified navigation styles are relevant for predicting disorientation. Most likely, other styles will be identified as well that can explain other facets of disorientation.

5.1 Implications for Adaptive Navigation Support

Prediction of users experiencing disorientation that is based on navigation measures has important practical consequences. From a usability point of view it is useful to identify those users who are at risk of experiencing disorientation and to assist them by adequate, and possibly personalized, navigation support.

Context information is important for effective navigation, as each navigation process is inextricably tied to the structure of the site. Two types of user context can be distinguished: the *structural context* and the *temporal context* [12]. Structural navigation aids – such as site maps, menus and index pages – describe a user's current location and navigation options; temporal navigation aids – such as the browser's back button, bookmarks and visual navigation histories – describe the way that led to this position.

Users that navigate in a *flimsy* manner appear not to be able to reconstruct their navigation paths and therefore are prone to get stuck. Visual navigation histories might help them out. In contrast, users that do not navigate *laboriously* enough and yet do not effectively exploit the site structure, can better be presented local or global site maps or a list of links to index pages. As these types of add-on navigation support typically consume a large amount of screen estate, it is desirable both for users with tools that they do not need.

5.2 Future Perspectives

In this paper we discussed how to address two navigation styles that might indicate or that might lead to users getting disoriented in web sites while working on goal-oriented tasks. The add-on navigation support, as discussed in the previous subsection, aims at improving the way users navigate rather than at forcing users to passively follow some ready-made paths. We believe that this should be the goal of adaptive hypermedia systems in general. Whereas the results of this pilot study might be applicable only in the small domain of web-assisted personal finance, the prospect of adaptive navigation support that fits the user's navigation style is attractive. As an example, users that prefer to extensively explore the sites that they visit, should be supported in doing so, instead of being urged to leave for a different site, unless the system is capable of making clear to the users that the benefit is higher than the cost of altering their strategy.

Acknowledgements. We would like to thank Herre van Oostendorp and Betsy van Dijk for their careful reviews of this paper. We also would like to express our

gratitude to Mark Neerincx from TNO Human Factors for providing us with standardized cognitive tests. Part of the research presented takes place in the context of the PALS Anywhere project, which is sponsored by the Dutch Innovative Research Program IOP-MMI

References

1. Ahuja, J.S. and Webster, J. Perceived disorientation: an examination of a new measure to assess web design effectiveness. *Interacting with Computers*, 14(1) (2001), 15-29.
2. Berendt, B. and Bernstein, E. Visualizing individual differences in Web navigation: STRATDYN, a tool for analyzing navigation patterns. *Behavior Research Methods, Instruments & Computers* 33 (2) (2001), 243-257.
3. Broder, A., Kumar, R., Farzin, M., Raghaven, P., Rajagopalan, S., Stata, R., Tomkins, A. & Wiener, J. Graph structure in the web. *Proc of 9th WWW Conference* (2000), 247-256.
4. Chen, C. Individual Differences in a Spatial-Semantic Virtual Environment. *Journal of The American Society for Information Science* 51 (6) (2000), 529-542.
5. Chen, S.Y. and Macredie, R.D. Cognitive Styles and Hypermedia Navigation: Development of a Learning Model. *Journal of the American Society for Information Science and Technology* 53 (1) (2002), 3-15
6. Cohen, J. A Power Primer. *Psychological Bulletin* 112 (1) (1992), 155-159.
7. Eveland, W.P. Jr. and Dunwoody, S. Users and navigation patterns of a science World Wide Web Site for the public. *Public Understanding of Science* 7 (1998), 285-311.
8. Herder, E. & Van Dijk, B. Site Structure and User Navigation: Models, Measures and Methods. In *Adaptable and Adaptive Hypermedia Systems* (ed. Chen, S.Y. & Magoulas, G.D.), to appear.
9. Juvina, I., and Van Oostendorp, H. Individual Differences and Behavioral Aspects Involved in Modeling Web Navigation. *Proc. 8th ERCIM Workshop 'User Interfaces For All'* (2004).
10. MacGregor, S., Hypermedia Navigation Profiles: Cognitive Characteristics and Information Processing Strategies. *Journal of Educational Computing Research* 20 (2) (1999), 189-206.
11. McEneaney, J.E. Graphic and numerical methods to assess navigation in hypertext. *Intl. J. Human-Computer Studies*, 55 (2001), 761-786.
12. Park, J. and Kim, J. Contextual Navigation Aids for Two World Wide Web Systems. *Intl. J. Human-Computer Interaction* 12 (2) (2000), 193-217.
13. Pirolli, P., and Card, S.K. Information foraging. *Psychological Review* 105(1) (1999) 58-82.
14. Rauterberg, M.: A Method of Quantitative Measurement of Cognitive Complexity. *Human-Computer Interaction: Tasks and Organisation* (1992), 295-307.
15. Rozanski, H. D., Bollman G. & Lipman M. Seize the Occasion: Usage-Based Segmentation for Internet Marketers. *eInsights – White Paper*, Booz-Allen & Hamilton (2001)
16. Shahabi, C., Zarkesh, A., Adibi, J. & Shah, V. Knowledge Discovery from Users Web-page Navigation. *Proc. of 7th Int. Workshop on Research Issues in Data Eng. on High Performance Database Management for Large-Scale Applications* (1997).
17. Tauscher, L. & Greenberg, S. How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies* 47 (1997), 97-137.
18. Thüring, M., Hanneman, J. & Haake, J.M. Hypermedia and Cognition: Designing for Comprehension. *Communications of the ACM* 38 (8) (1995), 57-66.

Designing an Adaptive Feedback Scheme to Support Reflection in Concept Mapping

Evangelia Gouli, Agoritsa Gogoulou, Kyparisia Papanikolaou, and
Maria Grigoriadou

Department of Informatics and Telecommunications, University of Athens,
Panepistimiopolis, GR-15784 Athens, Greece
{lilag, rgog, spap, gregor}@di.uoa.gr

Abstract. In this paper, we present an adaptive feedback scheme, which is incorporated in the “Knowledge Reconstruction + Refinement” process of a web-based concept mapping tool, named COMPASS, in order to support the reflection process in concept mapping. The feedback scheme includes multiple informative and tutoring feedback components and combines a stepwise presentation of these components with a multiple try strategy, aiming to provide personalized feedback. The adaptation of the scheme is based on the learner’s knowledge level, preferences and interaction behaviour. Two pilot empirical studies were conducted in order to investigate whether the design of the feedback components as well as the proposed adaptive feedback scheme can stimulate learners to reflect on their beliefs and appropriately revise their maps. The results revealed from the studies are encouraging, as the feedback provided, led the majority of the students to reconstruct/refine their knowledge and accomplish successfully the concept mapping tasks.

1 Introduction

Concept mapping, as a knowledge elicitation technique, stimulates learners to articulate and externalise their actual states of knowledge during the learning process. A concept map is comprised of nodes (concepts) and links (relationships between concepts), organized in a hierarchical structure to reflect the central concept of the map. Meaningful relationships between concepts form propositions. It is important to emphasize the inherently reflective nature of concept mapping, as it requires from learners to reflect on their understanding of concepts and their relationships [9].

Various applications of concept maps in learning and assessment and a number of concept mapping software tools are presented in [1]. During the assessment process, feedback is usually provided to learners according to specific common errors identified on their concept maps [2], [3]. These approaches do not take into account any learner’s individual characteristics or needs. More specifically, in [2], the system analyses the learner’s map by comparing it with the teacher’s map and provides hints (feedback strings defined by the teacher) about specific errors such as missing propo-

sitions. In [3], the system gives appropriate hints to the learner in the form of partial propositions. Moreover, to our knowledge, very few studies focus on the adaptation of the provided feedback according to learners' individual differences. In [7], a study was conducted, examining the effects of adaptive feedback (adjusting the amount of feedback based on learners' confidence in their answer) on learning outcomes and learning efficiency. In [6], a framework for the provision of feedback, based on the nature of the learning task and the learner's achievement level and prior knowledge, is presented. In [11], the incorporation of adaptive feedback into the proposed system is one of the researchers' plans.

In this context, we are developing a tool, named COMPASS (COncept MaP AS-Sessment tool) [4], aiming to provide a more flexible and learner-centered approach in the accomplishment of assessment activities based on concept mapping tasks and help learners to reconstruct/refine their knowledge. COMPASS supports the "Knowledge Reconstruction + Refinement" (KR+R) process by providing multiple informative and tutoring feedback components, tailored to the learners' knowledge level, preferences and interaction behaviour, through a stepwise presentation. The provided feedback aims to stimulate learners to reflect on their beliefs and proceed with the appropriate revisions. Two pilot empirical studies were conducted in order to investigate whether the design of the feedback components and the proposed adaptation scheme can help learners in revising their beliefs and refining their knowledge.

The paper is organized as follows. In Section 2, a description of the functionality of COMPASS is outlined. Then, in section 3, the adaptive feedback scheme, incorporated into the "KR+R" process, is presented. The results revealed by the two empirical studies are presented in Section 4 and the paper ends with concluding remarks and some directions for future work.

2 The COMPASS Tool

COMPASS is a web-based concept mapping tool aiming to assess the learner's understanding as well as to support the learning process. In particular, COMPASS serves (i) the *assessment process* by employing a variety of activities and applying a scheme for the qualitative and quantitative estimation of the learner's knowledge, and (ii) the *learning process* by providing different informative and tutoring feedback components, tailored to each individual learner, through the "KR+R" process.

More specifically, COMPASS supports the elaboration of assessment activities employing various mapping tasks such as the construction of a concept map from scratch ("free construction" task), the completion and evaluation of a concept map using an available list of concepts/relationships ("concept-relationship list completion/evaluation" task) [4]. After the learner has completed the assessment activity, COMPASS activates the *diagnosis process* for (i) the identification of errors on the learner's map (according to Table 1), based on the similarity of the learner's map to the teacher's one, and the qualitative analysis of the errors, (ii) the qualitative diagnosis of learner's knowledge, which is based on the proposed error categorization (Ta-

ble 1) and concerns the identification of the unknown concepts, incomplete understanding and false beliefs, and (iii) the quantitative estimation of learner's knowledge level on the central concept of the map and subsequently on the assessment activity, which is assigned to one of the characterizations {Insufficient (Ins), Rather Insufficient (RIns), Average (Ave), Rather Sufficient (RSuf) and Sufficient (Suf)}; this assignment is based on specific assessment criteria defined by teacher [4]. The learner may check/verify his/her map through the "Analysis" tool (Fig. 1). This tool provides the "Visual Feedback" option and the "Interactive Feedback" option. In case learner selects the "Visual Feedback" option, COMPASS graphically annotates the errors on the map, if any, following the proposed error categorization. In case of the "Interactive Feedback" option, COMPASS activates the "*KR+R*" process resulting to the provision of the appropriate feedback for each of the errors identified on the map.

3 The Adaptive Feedback Scheme

Feedback is considered as one of the most important sources of information to assist learners in restructuring their knowledge [6]. According to [5], effective feedback provides the learner with two types of information: verification (a judgement of whether the learner's answer is correct/incorrect) and elaboration (relevant cues to guide the learner toward a correct answer). Depending on the levels of verification and elaboration incorporated into the feedback, different types and forms of information may be combined (e.g. explanations for correct/incorrect answers, hints about useful sources of information, the knowledge of response) [6]. As one of the factors that contribute to the informative and tutoring value of feedback is the individual characteristics of the learner (e.g. learning objectives, prior knowledge and skills, motivational prerequisites), many researchers propose to tailor feedback to learner's individual needs and characteristics [10], [8].

In the context of COMPASS, the "*KR+R*" process aims to provide feedback, tailored to each individual learner in order to support the reflection process, to tutor and guide the learners and subsequently to enable them enrich/reconstruct their knowledge structure. The feedback scheme, adopted in the "*KR+R*" process, incorporates informative and tutoring feedback components (ITFC) and combines a stepwise presentation of these components with a multiple try strategy (see Activating the "*KR+R*" process). The ITFC include (i) an initiating question (IQ) consisting of the learner's belief, and a prompt to think of the concepts included in the proposition and to write any keywords describing the concepts, (ii) specific error-task related questions (E-TRQ), (iii) tutoring feedback units (TFU) relevant to concepts/relationship included in the concept map, and (iv) the knowledge of correct response (KCR). The ITFC concerning the E-TRQ and/or the TFU are provided according to the learner's individual characteristics (i.e. learners' knowledge level, preferences and interaction behaviour). Moreover, the stepwise presentation of the ITFC provides gradually the appropriate feedback components that are considered to be necessary in order the learner to modify/enrich his/her knowledge structure. Below, we present the design of

the E-TRQ and the TFU, the adaptation of the feedback scheme as far as these specific feedback components are concerned, and the stepwise feedback presentation.

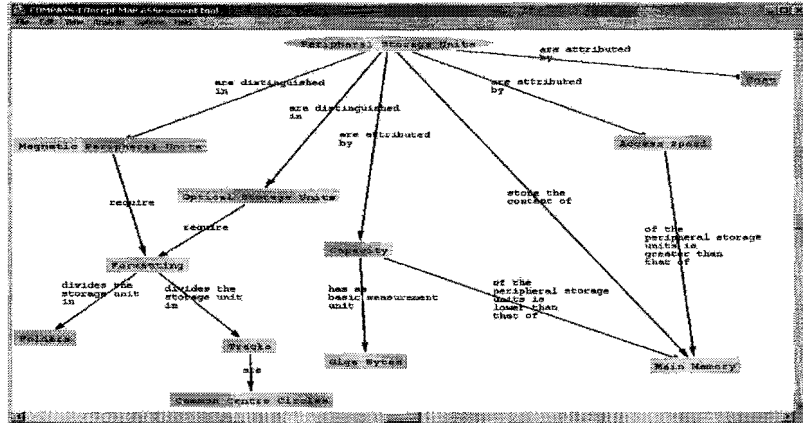


Fig. 1. A concept map constructed by a learner for a “concept-relationship list completion/evaluation” task.

The Design of the E-TRQ and the TFU. The error-task related questions, incorporated (E-TRQ) into the feedback scheme, aim to redirect the learner’s thinking and give a hint for correcting the error and completing the task. In the context of COMPASS, the form of the questions is differentiated according to the error categories that may be identified on the learner’s map. The form of the questions that are associated with each error category as well as an example of such a question for the learner’s map illustrated in Fig. 1, are presented in Table 1.

The tutoring feedback units (TFU) aim to allow the learner to review educational material relevant to the attributes of the desired/correct response. In the context of COMPASS, the TFU concern: (i) the concepts represented on the teacher’s concept map and/or the concepts included in the provided list of concepts (if a list of concepts is provided according to the mapping task) (TFUC), and (ii) specific propositions that the teacher anticipates a learner’s false belief (TFUP) [4]. TFUC are organised in two levels, TFUC1 and TFUC2 differing on the level of detail of the feedback information. TFUC1 presents the corresponding concept in general and it is independent of the mapping task (i.e. the same TFUC1 can be provided for different mapping tasks, which include the specific concept). TFUC2 presents the corresponding concept in more detail, focusing on the relationships of the concept with the other concepts of the map. Thus, TFUC2 depends on the concepts that may be represented on the particular concept map. TFUC2 is provided only if the learner insists on his/her belief after providing TFUC1. The feedback units (TFUC1 and TFUP) are associated with educational material consisting of knowledge modules, which constitute multiple representations of the concepts included in the proposition (i.e. a definition/description, an example, and/or an image of the concepts).

Table 1. The qualitative diagnosis of learners' knowledge based on different categories of errors and the form of error-task related questions according to the error categorization.

Qualitative Diagnosis of Learners' Knowledge	Categories of the Learners' Errors	
	Form of E-TQR [C1], [C2], [C3] ... are concepts, while R, R ₁ , R ₂ ... are relationships between concepts	Example of Error-Task Related Question
Unknown Concepts	Missing concept and its relationships: specific concepts, which should be represented on a map and have been defined by the teacher as fundamental concepts for the specific task/map [4], are missing.	
	<i>Do you consider that you could add on your concept map the concept of [C1]?</i>	Do you consider that you could add on your concept map the concept of [Sectors]?
Incomplete Understanding	Incomplete relationship: the relationships between two concepts are incomplete, as several relationships are missing (e.g. concepts [C1] and [C2] are related with m relationships on the teacher's map, while on the learner's map n relationships appear, where n<m).	
	<i>Do the [C1] only R the [C2]?</i>	Not applicable to the example
	Missing relationship: the relationship between two concepts that should be related is missing.	
	<i>Do you consider that you could add a relationship between the concepts of [C1] and [C2]?</i>	Do you consider that you could add a relationship between the concepts of [Cost] and [Main Memory]?
False beliefs	Superfluous relationship: two concepts are related although they should not.	
	<i>Do you really believe that the concepts [C1] and [C2] are related with the specific relationship?</i>	Do you really believe that the concepts [Optical Storage Units] and [Formatting] are related with the specific relationship?
	Incorrect relationship: two concepts are related with an incorrect relationship, which should be substituted.	
	<i>The [C1] R [C2]. Do you agree with this?</i> (where R is the correct relationship as represented on the teacher's concept map)	The [Capacity] of the peripheral storage units is greater than that of [Main Memory]. Do you agree with this?
	Superfluous concept: a superfluous concept appears which should be deleted.	
	<i>Do you want to reconsider the relationship of the concept [C2] with (i) the concept [C1], and (ii) the central concept of the map?</i>	Do you want to reconsider the relationship of the concept [Folders] with (i) the concept [Formatting], and (ii) the central concept of the map?
	Incomplete propositions: a concept (presented on the map) is not related to all the required concepts because the related concepts are missing.	
	<i>Do you really believe that [C1] only (i) R₁ [C2] and (ii) R₂ [C3]?</i>	Not applicable to the example
	Incorrect concept: a concept is related to an incorrect concept, which should be replaced with another concept.	
	<i>Do you really believe that the [C1] (correct concept) R [C2] (incorrect concept) [and the concept of [C2] is related with the concepts of [C3] and [C4]]?</i> (where the concepts [C3] and [C4] are the children of the concept [C2])	Do you really believe that the [Capacity] has as basic measurement unit [Gigabytes]*? *: the concept [Gigabytes] has not children, so the rest of the question is not applicable

The Adaptation of the Feedback Scheme. The adaptation of the feedback scheme, regarding the provision of TFU and/or E-TRQ, is based on information concerning the learner's knowledge level, preferences (i.e. preferences on ITFC and on knowledge modules) and interaction behaviour (i.e. knowledge modules of TFUC1 or TFUP more often provided, ITFC more often provided and frequency of errors made) (this information is provided by the learner model). Indicative rules that have been adopted in the adaptation scheme are:

- If the knowledge level of the learner has been evaluated as (Ins) or (RIns) on the assessment activity, then both TFU and E-TRQ are provided (TFU+E-TRQ).
- If the knowledge level of the learner has been evaluated as (Suf) or (RSuf) on the assessment activity, then E-TRQ are provided.
- If the knowledge level of the learner has been evaluated as (Ave) on the assessment activity, then according to the learner's preferences (ITFC preferred) and interaction behaviour (ITFC more often provided and frequency of errors made), E-TRQ or TFU+E-TRQ is provided. For example, (i) if the learner's favourite ITFC is E-TRQ but TFU+E-TRQ is more often provided, then TFU+E-TRQ is provided, (ii) if the frequency of a specific error identified on the learner's map is minimal (e.g. the learner's map includes very few incorrect relationships), then E-TRQ is provided.
- If TFU+E-TRQ is to be provided, then according to the error category, TFUC1 and/or TFUP is provided. TFUP is provided when the error belongs to the categories of "incorrect relationship", "incomplete relationship", "incomplete propositions" and "superfluous relationship". TFUC1 may concern more than one concepts according to the error category (e.g. in case of "incorrect concept", TFUC1 concerns only the incorrect concept of the proposition, while in case of "superfluous relationship", TFUC1 concerns both the concepts [C1] and [C2]).
- If TFUP and/or TFUC1 is to be provided and both types are available for the specific error (e.g. "superfluous relationship"), then TFUP is firstly provided and if the learner insists on his belief and/or asks for more help, TFUC1 is provided.
- If TFUC1 or TFUP is to be provided, then according to learner's preferences on knowledge modules and/or learner's interaction behaviour (types of knowledge modules more often provided), specific types of knowledge modules (i.e. definition/description, example and/or image of the concept) are provided.
- If the learner insists on his/her belief although TFUC1 was provided, then TFUC2 is also provided (in case it is available).

Activating the "KR+R" Process. COMPASS incorporates the abovementioned feedback scheme as well as the adaptation mechanism in the "KR+R" process. The "KR+R" process is activated when the learner completes an activity or asks for support/help during the task. The following sequence of interactions is taken place:

- *First Step:* After detecting an error on the learner's concept map, COMPASS indicates the error by providing the learner with an initiating question (IQ). The IQ gives learners the possibility to rethink their beliefs and to identify and check their own errors. This form of feedback may be sufficient for learners with high knowledge level. The applicability of the step depends on the category of error

(e.g. for a “missing relationship” error, this step is not applied). Following, the tool enters in a “wait” state, expecting the learner’s action.

- *Second Step:* If the learner insists on his/her belief, then according to the abovementioned rules E-TRQ and/or TFU+E-TRQ are provided. COMPASS enables the learner to think about the feedback and proceed with any changes; the tool enters again in a “wait” state, expecting the learner’s action.
- *Third Step:* If an impasse is reached (learner insists on his/her belief) or the learner asks for the knowledge of correct response, then COMPASS informs the learner about the correct response (KCR feedback component).

It is important to mention that during the interaction between the learner and the tool, the learner has always the option to select the feedback component and the knowledge modules that s/he prefers, ignoring the ones provided by the tool.

4 The Empirical Studies

The design of the “KR+R” process was carried out in parallel to two empirical studies that we conducted as a pilot evaluation before proceeding with the implementation of the process in the context of COMPASS. The two studies were carried out during the winter semester of the academic year 2003-2004, in order to investigate whether the design of the feedback components, as well as the adopted adaptation scheme, could stimulate learners to reflect on their beliefs and appropriately revise their maps.

First Empirical Study. In order to investigate whether the design of the E-TRQ, as the only source of feedback, can help learners towards the direction of identifying their errors, reconsidering and correcting them appropriately, we conducted an empirical study. Six high school students volunteered to take part. The students had to accomplish a “concept-relationship evaluation” concept mapping task concerning the central concept of “Magnetic Peripheral Storage Units”. After the accomplishment of the activity, the teacher interacted with each one of the students, simulating the step-wise presentation of the “KR+R”. The duration of the empirical study was 2 hours.

For the six students, the percentage of correct responses for each error category, before the provision of feedback and after the stepwise feedback presentation (for the 2nd step only the E-TRQ were provided), is presented in Fig. 2. The reader may notice that all the students improved their performance and the questions helped them to reconsider their beliefs and correct the majority of the errors. However, there are some cases that the questions didn’t help the students to find all the errors (e.g. the case of the 3rd student in the error categories of “incorrect relationship”, “superfluous relationship” and “missing relationship”). As far as any modifications to the form of the E-TRQ are concerned, the study drew implications about the form of the questions posed for the error categories of “incorrect concept” and “superfluous relationship” (the modified versions of the questions are presented in Table 1). Regarding the process (i.e. steps), it is important to mention that in several cases, the application of the first step (i.e. the provision of the IQ) was proved to be adequate and helped students to check for accidental constructions. There were cases that the students weren’t

able to correct their errors even if E-TRQ were provided; in these cases, the teacher tried to explain in details the concepts involved in the proposition. This observation led us draw the conclusion that the specific ITFC (i.e. IQ and E-TRQ) are not adequate in all cases; additional feedback should be provided.

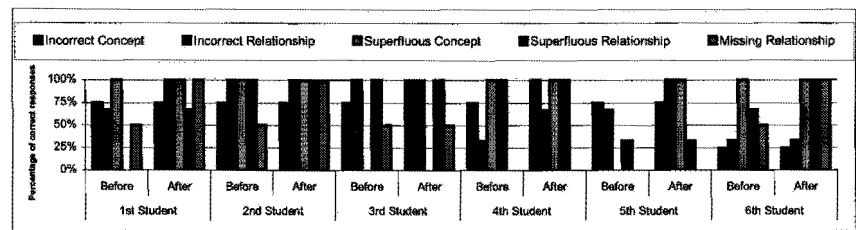


Fig. 2. The percentage of the correct responses concerning specific categories of errors.

Summarizing the results, it seems that the form of the E-TRQ can help students, especially those with knowledge level above average, in revising their beliefs and refining their knowledge. In cases of students with low knowledge level, a form of tutoring feedback is required in order to help them identify and correct their beliefs. Therefore it was considered important to incorporate TFU in the feedback scheme.

Second Empirical Study. In the second empirical study, the feedback provided to the learners included both the TFU and the E-TRQ. The aim of this study was to investigate whether the design of the proposed adaptation scheme, can stimulate learners to reflect on their beliefs and appropriately revise their maps. Ten high school students volunteered to take part in the study, which lasted 3 hours.

A pre-test was conducted in order to estimate the students' prior knowledge level. The pre-test had the form of open questions such as "Mention keywords that describe the concept of Formatting", "Mention the kinds of Peripheral Storage Units". The pre-test questions address the concepts/relationships that could be represented on the map of the task that the students had to accomplish after the pre-test. The teacher assessed their answers and estimated their knowledge level (1 student as (Suf), 3 as (RSuf), 3 as (Ave), 2 as (RIns), and 1 as (Ins)). The students' preferences concerning the types of knowledge modules (description, example or image) and the ITFC (TFU+E-TRQ and E-TRQ) were also recorded. The task, that the students had to accomplish, was a "concept-relationship list completion/evaluation" task. After its accomplishment, the teacher interacted with each one of the students, simulating the stepwise presentation of the "KR+R". To this end, the learner's interaction behaviour was not considered.

The 1st step of the process (i.e. the IQ feedback component) was adequate only for one student (the 3rd student claimed that he made the errors by accident and was able to recognize and correct them). In the context of the 2nd step of the process, the E-TRQ were used for those students whose knowledge level was characterized as (Suf) and (RSuf). The E-TRQ were proved to be effective in helping the students to identify their errors and correct them appropriately (see Fig. 3). All the students whose knowledge level was characterized as (RIns) and (Ins), improved their performance

(see Fig. 3) after the TFU+E-TRQ were provided and they identified and corrected a considerable number of errors. Two of them (5th and 6th student) didn't manage to correct all the errors; in two error cases the KCR was finally provided. In the case of the 7th student, the TFUC1 and TFUC2 were provided, helping him to accomplish correctly the mapping task. For those students whose knowledge level was characterized as (Ave), their preferences concerning the TFU+E-TRQ and E-TRQ (one student selected E-TRQ and two students selected TFU+E-TRQ) were taken into account. All three students, after the provision of feedback, accomplished the task successfully.

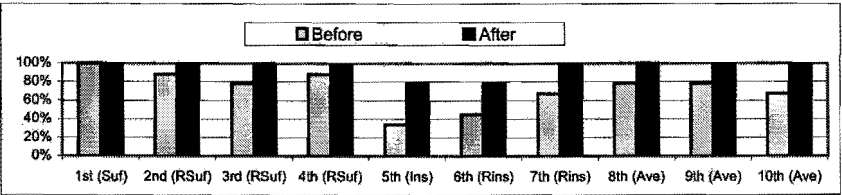


Fig. 3. The percentage of correct responses before and after the provision of feedback

Summarizing the results, it seems that the ITFC that were provided, following the stepwise presentation stimulated students to review their maps and reconsider their beliefs, as the majority of them spent some time thinking of them. It has to be mentioned, that the teacher, in all the cases, tried to elicit from students why they proceed with the desired corrections. The impression was that the students had fully understood their errors and refined their knowledge. The adaptive feedback scheme can be characterized as promising as the majority of the students accomplished successfully the mapping task and refined their knowledge. The results revealed from the two studies provided useful indications on the effectiveness of the proposed adaptive feedback scheme. However, data gathered from a larger sample, using COMPASS in real working conditions, under longer periods of time, are considered necessary for the aim of inferring learners' attitudes and evaluating the effectiveness of the adaptive feedback scheme.

5 Conclusions and Further Research

In this paper, we presented an adaptive feedback scheme, which is incorporated in the "Knowledge Reconstruction + Refinement" (KR+R) process of COMPASS in order to support the reflection process in concept mapping tasks. The discriminative characteristics of the "KR+R", and in particular of the proposed adaptation scheme are: the adoption of different informative and tutoring feedback components (ITFC) and the stepwise feedback presentation, the adoption of error-task related questions (E-TRQ) based on a categorization of learners' common errors, the adoption of the two levels of the tutoring feedback units (TFU) and the adaptation of feedback to the learner's knowledge level, preferences and interaction behaviour. The results from two empiri-

cal studies conducted, even performed on a limited number of subjects and in a simulated environment, are encouraging indicating that the provided feedback support reflection and help students to identify and correct their errors.

The presented research work contributes to the field of adaptive feedback, giving some promising directions for further research. Additional studies need to be conducted in order to compare the efficiency of the proposed informative and tutoring feedback components to other feedback components such as the knowledge of response and the effects of the proposed adaptive scheme to a standard feedback scheme as it is implemented in most learning environments. Our future plans include the enrichment of the informative and tutoring feedback components with additional forms as well as the conduction of a series of empirical studies with a wider group of learners, in order to evaluate COMPASS regarding the effectiveness of the provided feedback components and the adaptive feedback scheme.

References

1. Cañas, A., Coffey, J., Carnot, M., Feltovich, P., Hoffman, R., Feltovich, J., Novak, J.: A Summary of Literature Pertaining to the Use of Concept Mapping Techniques and Technologies for Education and Performance Support. Final report to CNET. Retrieved 2003 from <http://www.ihmc.us/users/acanas/Publications/ConceptMapLitReview/IHMC%20Literature%20Review%20on%20Concept%20Mapping.pdf>
2. Cimolino, L., Kay, J., Miller, A.: Incremental student modelling and reflection by verified concept-mapping. In Supplementary Proceedings of the AIED2003: Learner Modelling for Reflection Workshop (2003) 219-227
3. Chang, K., Sung, T., Chen, S-F.: Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning* 17 (1) (2001) 21-33
4. Gouli, E., Gogoulou, A., Papanikolaou, K., Grigoriadou, M.: COMPASS : An Adaptive Web-based Concept Map Assessment Tool. In Proceedings of the First International Conference on Concept Mapping (CMC 2004) (2004) (to appear)
5. Kulhavy, R Stock, W.: Feedback in written Instruction: The place of response certitude. *Educational Psychology Review* 1(4) (1989) 279-308
6. Mason, B., Bruning, R.: Providing Feedback in Computer-based Instruction: What the research tells us. Retrieved 2004 from <http://dwb.unl.edu/Edit/MB/MasonBruning.html>
7. Mory, E.: Adaptive feedback in computer-based instruction: Effects of response certitude on performance, feedback-study time, and efficiency. *Journal of Educational Computing Research* 11 (3) (1994) 263-290
8. Narciss, S., Huth, K.: How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, R. Brünken & D. Leutner (Eds.) *Instructional design for multimedia learning*. Münster, Germany: Waxmann (2002) 181-195
9. Novak, J., Gowin, D.: *Learning how to learn*. Cambridge University Press, UK (1984)
10. Schimmel, B.: Providing meaningful feedback in courseware. In D. Jonassen (Ed.), *Instructional designs for microcomputer courseware*. Hillsdale, NJ: LEA (1988) 183-195
11. Tsai, C-C., Chou, C.: Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted learning* 18 (2002) 157-165

Adaptation Rules Relating Learning Styles Research and Learning Objects Meta-data

Charalampos Karagiannidis^{1,2} and Demetrios Sampson^{2,3}

¹ University of the Aegean

Department of Cultural Technology and Communication

Mytilini, Lesvos Island, Greece

E-mail: karagian@acm.org

² Centre for Research and Technology - Hellas (CE.R.T.H.)

Informatics and Telematics Institute (I.T.I.)

Thessaloniki, Greece

<http://www.iti.gr>

³ University of Piraeus

Department of Technology Education and Digital Systems

Piraeus, Greece

E-mail: sampson@iti.gr

Abstract. This paper investigates the development of adaptation rules which relate individual learning styles characteristics to learning objects characteristics, as the latter are reflected in the IMS Learning Resource Meta-Data Specification. The paper outlines the most well-known learning styles theories and models, some criteria for selecting among them, as well as a number of adaptive web-based learning environments which are utilising learning styles research for facilitating personalised learning. The paper concludes with the set of adaptation rules which are used in the KOD learning environment, which aims to facilitate individualised access to learning material in a reusable way.

1. Introduction

The rapid evolution of Information and Communication Technologies (ICT) and the emergence of the Knowledge Society create numerous new opportunities for the improvement of the quality of education. It can be argued, however, that education has not yet realised the full potential of the employment of ICT: *“there is a shortage of solid evidence to back up the belief that telematic learning systems provide real advantages”* [1]. Besides the apparent benefits of ICT for delivering education to distance learners independently of time, location, etc, several studies question whether there is *“a significant difference”* with respect to the learning effectiveness when ICT is employed in education [2]. This is mainly due to the fact that the “traditional” mode of instruction (one-to-many lecturing, or one-to-one tutoring) adopted in “conventional” educational technology cannot not fully accommodate the different learning and studying styles, strategies and preferences of diverse learners.

Personalised learning (PL) systems are attracting increasing interest in this context, since they bare the potential to meet the requirements of the knowledge society and knowledge-based economy for high-quality education and training [3]. PL systems can be defined by their capability to automatically *adapt* to the changing attributes of

the “learning experience”, which can, in turn, be defined by the individual learner characteristics, the type of the learning material, etc. That is, PL systems can be categorised and differentiated in terms of their *adaptation logic*, which is defined by: (i) PL constituents: the aspects of the learning experience which are subject to adaptations; that is, is the learning material being adapted? and if so, how do we categorise learning material so that we can select different content for different learner? (ii) PL determinants: the aspects of the learning experience which “drive” adaptations; that is, are adaptations based on the learner’s profile? and if so, how is the learner profile defined? and (iii) PL rules: the rules which define which PL constituents are selected for different PL determinants [4]. PL systems can be quite diversified according to their adaptation logics, depending on the requirements of the specific learning context. For example, PL determinants can include learners’ characteristics, which can, in turn, include learner’s background, expertise, prior knowledge, skills, requirements, preferences, etc [5].

This paper addresses the incorporation of learning styles research in the adaptation logic of PL systems. That is, the definition of PL determinants, constituents and rules which are based on, and reflect specific learning styles theories and models. The next section provides a short overview of the most well-known learning styles theories and models, as well as some criteria for selecting among them when developing PL systems. Subsequently, the paper outlines some existing PL systems which utilise learning styles research, with emphasis on PL system which has been developed in the context of the KOD “Knowledge on Demand” European project (see acknowledgements section). The paper concludes with the set of adaptation rules used in the KOD project, which attempt to relate individual learning styles characteristics (as adaptation determinants), and learning objects characteristics (as adaptation constituents), as the latter are reflected in the IMS Learning Resource Meta-Data Specification – LOM [6].

2. A Brief Overview of Learning Styles

The term “learning styles” has been attributed with several connotations in the literature. Learning styles can be generally described as “an individual’s preferred approach to organising and presenting information” [7]; “the way in which learners perceive, process, store and recall attempts of learning” [8]; “distinctive behaviours which serve as indicators of how a person learns from and adapts to his environment, and provide clues as to how a person’s mind operates” [9]; “a gestalt combining internal and external operations derived from the individual’s neurobiology, personality and development, and reflected in learner behaviour [10].

Learning styles models can be presented through an onion metaphor (proposed in [11]), consisting of three basic layers which categorise learners in terms of instructional preferences (outermost layer), information processing (middle layer) and personality (innermost layer). Social interaction, a fourth layer placed between Curry’s two outer layers, was proposed in [12]. The most well-known and used learning styles theories and models are presented in Table 1. For each model, the presentation includes: (i) the learner categorisations proposed by each model, (ii) the existence of an assessment instrument for categorising each learner in the above categories, and (iii) indicative references for each model.

Table 1. Overview of Learning Styles

Name	Learners' Categorisation	Assessment Instrument
Kolb Learning Style Inventory [13], [14]	Divergers (concrete, reflective), Assimilators (abstract, reflective), Convergers (abstract/active), Accommodators (concrete/active)	Learning Style Inventory (LSI), consisting of 12 items in which subjects are asked to rank 12 sentences describing how they best learn.
Dunn and Dunn – Learning Style Assessment Instrument [15], [16]	Environmental, Emotional, Sociological, Physical factors.	(i) Learning Style Inventory (LSI) designed for children grade 3–12; (ii) Productivity Environmental Preference Survey (PEPS) – adult version of the LSI containing 100 items
Felder-Silverman – Index of Learning Styles [17], [18]	Sensing-intuitive, Visual-verbal, Indicative-deductive, Active-reflective, Sequential-global	Soloman and Felder questionnaire, consisting of 44 questions
Riding – Cognitive Style Analysis [19], [20]	Wholists-Analytics, Verbalisers-Imagers	CSA (Cognitive Styles Analysis) test, consisting of three sub tests based on the comparison of the response time to different items
Honey and Mumford – Learning Styles Questionnaire [21]	Theorist, Activist, Reflector, Pragmatist	Honey & Mumford's Learning Styles Questionnaire (LSQ), consisting of 80 items with true/false answers
Gregoric – Mind Styles and Gregoric Style Delineator [9], [22]	Abstract Sequential, Abstract Random, Concrete Sequential, Concrete Random	Gregoric Style Delineator containing 40 words arranged in 10 columns with 4 items each; the learner is asked to rank the words in terms of personal preference
McCarthy – 4 Mat System [23], [24]	Innovative, Analytic, Common sense, Dynamic	-
Gardner – Multiple Intelligence Inventory [25], [26]	Linguistic, Logical-mathematical, Musical, Bodily-kinesthetic, Spatial, Interpersonal, Intrapersonal	an instrument consisting of 8 questions
Grasha-Riechmann – Student Learning Style Scale [27], [28]	Competitive-Collaborative, Avoidant-Participant, Dependent-Independent.	90 items self-report inventory measuring the preferences of both high school and college students
Hermann – Brain Dominance Model [29], [30]	Quadrant A (left brain, cerebral), Quadrant B (left brain, limbic), Quadrant C (right brain, limbic), Quadrant D (right brain, cerebral)	120 questions that refer to four profile preferences codes corresponding to each quadrant
Mayers-Briggs – Type Indicator [31], [32]	Extroversion, Introversion, Sensing, Intuition, Thinking, Feeling, Judgement, Perception	(i) MBTI (Myers-Briggs Type Indicator), (ii) Kiersey Temperament Sorter I, and (iii) Kiersey Character Sorter II

2.1 Criteria for Selecting Among Different Learning Style Models in PL Systems

Given the variety of learning styles theories and models that are available in the literature, we need to define a set of criteria for selecting the most appropriate learning style model to be accommodated in a specific PL system.

Of course, the most important criterion, apart from the theoretical and empirical justification of the model, is the suitability of each model for the specific learning context under consideration, the available adaptation technologies, etc, especially from an educational point of view. For example, if all learners of a specific learning context are “experts” in the domain (e.g. an application for aircraft pilots), then it might not be reasonable to select a learning style model which categorises learners as being either “experts” or “novices”. Similarly, if all the educational material that is available for a specific case is in textual form, then it is not reasonable to select a model which differentiates content according to its medium.

The following paragraphs summarise some additional selection criteria that need to be considered in this context.

- *Measurability*: We need to be able to “measure” how learners are “classified” into the categories defined by each model. For example, one model may differentiate learners according to their emotions; while this may be reasonable from a theoretical point of view, since emotions may affect learning, it may not be reasonable to select such a model for a PL system, since it may be very difficult to measure learners’ emotions. The existence of an assessment instrument (e.g. such as the questionnaires included in Table 1) may help in this direction. Moreover, for adaptive learning environments, this classification needs to be performed at run-time, based on the learners’ observable behaviour (i.e. it cannot be based on initial questionnaires).
- *Time effectiveness*: The assessment instrument related to each learning style model needs to include a reasonable number of questions in order to be time effective. For example, if an assessment instrument consists of 200 questions, then the instrument may not be effective time wise. The user may not be willing to dedicate his/her time in order to complete a large questionnaire before starting using the system.
- *Cost*: The cost of a learning style model along with its assessment instrument is another parameter that system designers may need to consider. The situation here varies, as some assessment instruments are only available for use after payment, while others are available to be used free-of-charge. Another type of “cost” related to each learning style model is the type of learning material selected for different (categories of) learners; for example, starting from a text-based learning material, it may not be cost-effective to adopt a visual/verbal learner classification, since this may require that the learning material is enriched with visual multimedia components.

3. Some Examples of Accommodating Learning Styles Research in PL Systems

Learning styles research has formed the basis for the development of a number of PL systems, since a number of studies have shown that adaptation to the individual’s learning style can have a positive impact on learning effectiveness (e.g. [33]). TrainingPlace.com is a notable example of a commercial PL system which is based on learning styles research. This system is based on Learning Orientation Theory, which categorises learners into transforming, performing, conforming and resistant. Based on this categorisation, the system presents different “learning experiences” to each learner. For example, the system selects “loosely structured environments that promote challenging goals, discovery and self-managed learning” for transforming learners, and “semi-complex, semi-structured, coaching environments that stimulate personal value and provide creative interaction” for performing learners [34].

The INSPIRE system adapts the presentation of the learning material, based on the Honey and Munford’s learning styles model, as shown below [35]:

learner style	selected learning material
activist (motivated by experimentation and challenging tasks)	activity-oriented learning material with high interactivity level
reflector (tend to collect and analyse data before taking action)	example-oriented learning material
pragmatist (keen on trying out ideas, theories and techniques)	exercise-oriented learning material
theorist (preferring to explore and discover concepts through more abstract ways)	theory-oriented learning material

The same learning styles model is used in SMILE, a web-based knowledge support system aiming at promoting intelligent support for dealing with open-ended problem situations [36]; as well as within the 3DE European Project www.3deproject.com), where different courses are developed for each learner from a repository of learning objects.

The AES-CS system adapts the learning environment based on the field dependent/independent learning styles model, as shown below [37]:

field-dependent learners	field-independent learners
provide global approach	provide analytical approach
provide information from general to specific	provide information from specific to general
program control	learner control
provide advance organizer	provide post organizer
provide maximum instructions	provide minimal instructions
provide maximum feedback	provide minimal feedback
provide structured lessons	allow learners to develop their own structure

The *iWeaver* system adapts the presentation of the learning material, based on the learner’s style, following the Dunn & Dunn model, as shown below [38]:

learner style	recommended representation	representation type
auditory	PowerPoint-style presentations with synchronous audio, no text	multimedia representation
visual (pictures)	diagrams, illustrations, graphs, flowcharts, animations & audio	multimedia representation & text or audio
visual (text)	reading, context-aware note-taking tool	text & additional tool
tactile kinesthetic	interactive multimedia elements: puzzles, drag-and-drop fill-ins, small games	multimedia representation & text
internal kinesthetic	extra examples of real-life relevance, links to prior content	additional text
impulsive	try-it button (allows immediate trial)	additional tool
reflective	context-aware note-taking tool, questions that encourage reflection	additional tool
global	advance organisers or mind maps	additional multimedia representation
analytical	sequential lists of key points and components	text (default)

4. Adaptation Rules in the KOD Adaptive Web-Based Learning Environment

The KOD European project aims to promote individualised access to learning material in a re-usable way. This section focuses on the adaptation rules which are used in the KOD project, in order to accommodate learning styles research (the work of the project has been published in [39], [40], [41]).

Following the discussion of the previous section (concerning the selection among the available learning styles theories and models), we should also highlight another important selection criterion: it is important that the selected model describes not only how learners are categorised, but also how instruction should be adapted for each learner category; that it, apart from the descriptive information (e.g. learners are categorised into “active” and “reflective”), the model should provide prescriptive guidelines, which can lead to specific adaptation rules for designing instruction and adaptation (e.g. what types of educational content should be selected for active and reflective learners).

However, learning styles models are usually rather descriptive in nature, in the sense that they offer guidelines as to what methods to use to best attain a given goal; they are not usually prescriptive in the sense of spelling out in great detail exactly what must be done and allowing no variation: “prescription only applies to deterministic or positivistic theories, which are almost nonexistent in the social sciences” [42].

In this context, as part of the KOD project, we “interpreted” the literature on the respective models (presented in Table 1), in order to develop a set of adaptation rules, which are shown in Table 2. Since KOD aims to build on existing learning

technologies specifications, the adaptation constituents (i.e. learning objects characteristics)¹ were selected among the LOM elements.

It should be noted that:

- the values of some of the LOM elements included in Table 2 require an extension of the current version of the specification; for example, the value theoretical does not belong into the suggested values of the educational.learningResourceType element;
- the values for the technical.format element have been selected for presentation purposes to be visual, verbal, etc; according to the specification, these values should be mapped onto MIME types, based on RFC2048:1996 specification (e.g. image/jpeg or image/gif, etc);
- the adaptation determinants of the rules of Table 2 could also be described through learning technologies specifications, and in particular through the Learner Information Package Specification – LIP [43]; since the LIP specification does not include specific elements for maintaining the learner’s learning styles characteristics, the adaptation determinants could be maintained within the (extended) preference.cognitive element.

Table 2. Example KOD Adaptation Rules

Felder-Silberman Index of Learning Styles
<ul style="list-style-type: none">• IF learner=sensing THEN LOM.educational.learningResourceType=exercise OR simulation OR experiment• IF learner=intuitive THEN LOM.educational.learningResourceType=problemStatement OR narrativeText• IF learner=visual THEN LOM.technical.format=visual• IF learner=verbal THEN LOM.technical.format=verbal• IF learner=inductive THEN LOM.educational.semanticDensity=0 OR 1• IF learner=deductive THEN LOM.educational.semanticDensity=3 OR 4• IF learner=active THEN LOM.educational.semanticDensity=exercise OR simulation OR experiment• IF learner=reflective THEN LOM.educational.semanticDensity=problemStatement OR narrativeText• IF learner=sequential THEN LOM.educational.semanticDensity=0 OR 1• IF learner=global THEN LOM.educational.semanticDensity=2 OR 3
Riding Cognitive Style Analysis
<ul style="list-style-type: none">• IF learner=analytic THEN LOM.educational.semanticDensity=0 OR 1• IF learner=wholist THEN LOM.educational.semanticDensity=2 OR 3• IF learner=visual THEN LOM.technical.format=visual• IF learner=verbal THEN LOM.technical.format=verbal

¹ The term “learning object” is used to refer to “any (digital) entity that can be used, re-used, or referenced during technology-supported learning” [6].

Honey and Mumford Learning Styles

- IF learner=theorist THEN LOM.educational.learningResourceType=theoretical
- IF learner=activist THEN LOM.educational.learningResourceType=practical
- IF learner=reflector THEN LOM.educational.semanticDensity=0 OR 1
- IF learner=pragmatist THEN LOM.educational.semanticDensity=2 OR 3

Gregoric – Mind Styles and Style Delineator

- IF learner=abstractSequential THEN LOM.educational.semanticDensity=0 OR 1
- IF learner=abstractRandom THEN LOM.educational.semanticDensity=2 OR 3
- IF learner=concreteSequential THEN
LOM.educational.learningResourceType=exercise OR simulation OR experiment
- IF learner=concreteRandom THEN
LOM.educational.learningResourceType=problemStatement OR narrativeText

Learning Orientation Theory

- IF learner=transforming THEN LOM.educational.interactivityLevel>2
- IF learner=performing THEN LOM.educational.interactivityLevel>2
- IF learner=conforming THEN LOM.educational.semanticDensity<2

5. Conclusions and Future Work

This paper investigates the accommodation of learning styles research in PL systems. It briefly reviews the most well-known learning styles theories and models, as well as some criteria for selecting among them, and also outlines a number of PL systems which utilise this line of research for delivering personalised learning.

The emphasis of the paper is on the PL system which has been developed in the context of the KOD European project, and, moreover, on the adaptation rules which have been used in the KOD system, based on learning styles research.

As it was described earlier, the “rule templates” employed in KOD (shown in Table 2) are the result of an “interpretation” of the literature on the respective models, which was carried out within the project. Part of our current and future work involves the validation of these “prescriptive rules”, through the development and testing of adaptive learning material which is based on them. The KOD project, facilitating the interchange of learning material in a re-usable way [40], offers an effective test-bed for this endeavour.

Acknowledgements

Part of the R&D work reported in this paper was carried out in the context of the KOD “Knowledge on Demand” project (www.kodweb.org, kod.iti.gr), which was partially funded by the European Commission, under the Information Society Technologies Programme (Contract No IST-1999-12503). The KOD Consortium comprises: CERTH-ITI, Greece (project co-ordinator); FD Learning, UK; GUINTI Interactive Labs, Italy, CATAI, Spain; and PROFit Gestion Informatica, Spain.

References

1. European Commission, Future European RTD Agenda for Technology Enabled Learning, Report on the 1st Consultation Meeting, Luxembourg, March 23-24, 2000.
2. Russell T.L., The No Significant Difference Phenomenon as reported in 355 Research Reports, Summaries and Papers: A Comparative Research Annotated Bibliography on Technology for Distance Education, North Carolina State University, Office of Instructional Telecommunications, 1999.
3. Brusilovsky P., Adaptive and Intelligent Technologies for Web-based Education, *Kunstliche Intelligenz*, Special Issue on Intelligent Systems and Teleteaching, 4, 1999.
4. Stephanidis C., Karagiannidis C. and Koumpis A., Decision Making for Intelligent User Interfaces, 3rd ACM International Conference on Intelligent User Interfaces (IUI 97), Orlando, Florida, January 1997.
5. Brusilovsky P., Adaptive Hypermedia, *User Modeling and User-Adapted Interaction*, 11(1), 2001.
6. IMS Global Learning Consortium, Learning Resource Meta-Data Specification, Version 1.2.1, September 2001. Electronically available from <http://www.imsglobal.org/metadata/index.cfm>.
7. Riding R & Rayner S., *Cognitive styles and learning strategies: Understanding style differences in learning and behaviour*, David Fulton Publishers, 1998.
8. James W., & Gardner D., *Learning styles: Implications for distance learning*, New Directions for Adult and Continuing Education, 67, 1995
9. Gregoric A., *Learning/teaching styles: Potent forces behind them*, *Educational Leadership*, 36(4), 1979.
10. Keefe J., & Ferrell B., *Developing a Defensible Learning Style Paradigm*, *Educational Leadership*, 48 (2), 1990
11. Curry L., *An Organisation of Learning Styles Theory and Constructs*, American Educational Research Association Conference, Montreal, Quebec, 11-15 April, 1983.
12. Claxton C., & Murrell P., *Learning Styles: Implications for Improving Education Processes*, ASHE-ERIC Higher Education Report No 4, Association for the Study of Higher Education, 1987.
13. Kolb A., *Experiential Learning: Experience as the Source of Learning and Development*, Prentice Hall, 1984.
14. Kolb A., *LSI Learning-Style Inventory*, McBer & Company, 1985.
15. Dunn R., & Dunn K., *Teaching Students through their Individual Learning Styles: A Practical Approach*, Reston Publishing Company, 1978.
16. Dunn R., & Dunn K., *The Complete Guide to the Learning Styles Inservice System*, Allyn & Bacon, 1999.
17. Felder R., & Silverman L., *Learning and Teaching Styles in Engineering Education*, *Engineering Education*, 78(7), 1988.
18. Felder R., *Matters of Style*, *ASEE Prism*, 6(4), 1996.
19. Riding R., & Cheema I., *Cognitive Styles: An Overview and Integration*, *Educational Psychology*, 11(3&4), 1991.
20. Riding R., *Cognitive Styles Analysis*, *Learning and Training Technology*, 1994.
21. Honey P., & Mumford A., *The Manual of Learning Styles*, 3rd Ed., Maidenhead, Peter Honey, 1992.
22. Gregoric A., *Style Delineator*, Maynard MA, Gabriel Systems, 1982.
23. McCarthy B., *The 4MAT System: Teaching to Learning Styles with Right/Left Mode Techniques*, Excel. Inc, 1980.
24. McCarthy B., *A Tale of Four Learners: 4MAT's Learning Styles*, *Educational Leadership*, 1997.
25. Gardner H., *Frames of Mind: The Theory of Multiple Intelligences*, 2nd Ed., Fontana, 1993.

26. Gardner H., *Multiple Intelligences: The Theory in Practice*, Basic Books, 1993.
27. Hruska-Riechmann S., & Grasha A., *The Grasha-Riechmann Student Learning Style Scales: Research Findings and Applications*, in Keefe J. (ed.), *Student Learning Styles and Brain Behaviour*, Reston, 1982.
28. Grasha A., *Teaching with Style: A Practical Guide to Enhancing Learning by Understanding Teaching and Learning Styles*, Alliance Publishers, 1996.
29. Herrmann N., *Herman Brain Dominance Instrument*, Lake Lure, Applied Services, 1982.
30. Herrmann N., *The Creative Brain*, Lake Lure, Brain Books, 1995.
31. Myers I., & Kirby L., *Introduction to Type Dynamics and Development*, Consulting Psychologists Press, 1994.
32. Myers I., McCaulley M., Quenk N., & Hammer L., *The MBTI Manual: A guide to the Development and Use of the Myers-Briggs Type Indicator*, Consulting Psychologists Press, 1998.
33. Graff M., *Assessing Learning from Hypertext: An Individual Differences Perspective*, *Journal of Interactive Learning Research*, 14(4), 2003.
34. Martinez M., & Bunderson C., *Foundations for Personalised Web Learning Environments*, *ALN Magazine*, 4(2), 2000.
35. Papanikolaou K., Grigoriadou M., Kornilakis H., & Magoulas G., *Personalising the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE*, *User-Modeling and User-Adapted Interaction*, 13 (3), 2003.
36. Stoyanov S., & Kommers P., *Agent-Support for Problem Solving Through Concept-Mapping*, *Journal of Interactive Learning Research*, Special issue on Intelligent Agents from Educational Computer-Aided Systems, 10(3/4), 1999.
37. Triantafyllou E., Pombortsis A., & Demetriadis S., *The Design and the Formative Evaluation of an Adaptive Educational System Based on Learning Styles*, *Computers and Education*, 41, 2003.
38. Wolf C., *iWeaver: Towards an Interactive Web-Based Adaptive Learning Environment to Address Individual Learning Styles*, *European Journal of Open and Distance Learning*, 2002.
39. Karagiannidis C., Sampson D., & Cardinali F., *Integrating Adaptive Educational Content into Different Courses and Curricula*, *Educational Technology & Society Journal*, IEEE Learning Technology Task Force, Special Issue on "Curriculum, Instruction, Learning and the Internet", 4(3), July 2001.
40. Sampson D., Karagiannidis C., & Cardinali F., *An Architecture for Web-based e-Learning Promoting Re-usable Adaptive Educational e-Content*, *Educational Technology & Society Journal*, IEEE Learning Technology Task Force, Special Issue on Innovations in Learning Technology, 5(4), August 2002 (Invited Paper).
41. Sampson D., Karagiannidis C. & Kinshuk, *Personalised Learning: Educational, Technological and Standardisation Perspectives*, *Interactive Educational Multimedia*, 4, April 2002.
42. Reigeluth C.M., *The Elaboration Theory: Guidance for Scope and Sequence Decisions*, in Reigeluth C.M. (ed.), *Instructional-Design Theories and Models Volume II*, Lawrence Erlbaum Associates, p. 425-453, 1999.
43. IMS Global Learning Consortium, *Learner Information Package Specification, Version 1.0*, March 2001. Electronically available from <http://www.imsglobal.org/profiles/index.cfm>.

Adapting Hypermedia to Cognitive Styles: Is it necessary?

Timothy Mitchell, Sherry Y. Chen, and Robert Macredie

Department of Information Systems and Computing, Brunel University, Uxbridge,
Middlesex, UB8 3PH, UK
{Tim.Mitchell, Sherry.Chen, Robert.Macredie}@brunel.ac.uk

Abstract. Adaptive hypermedia systems traditionally focused on adapting to the user's prior knowledge, but recent research has begun to consider adapting to cognitive style. This paper presents the results of an experiment investigating the learning performance and user-perceptions of 60 undergraduate students using educational adaptive hypermedia interfaces. Participants used two interfaces – one 'normal' and one adaptive – and were randomly matched or mismatched to their cognitive styles. Whilst there was no interface preference for those who were matched to their cognitive styles, those who were mismatched were significantly more likely to prefer the normal interface. The implications of these findings in relation to adaptive hypermedia development are discussed.

1. Introduction

Adaptive hypermedia (AH) is hypermedia that can adapt the content presentation and navigation support to the users, to aid the users in their search for the information most appropriate and best suited to them [1]. The purpose of AH, particularly in regard to education, is to provide a learning environment that can match with the needs of each individual, preventing them from being "lost in hyperspace" [2]. A number of AH systems have been created for educational settings and these systems tailor information to the students' level of knowledge such as ELM-ART [3] and InterBook [7]. Such systems adapting to prior knowledge have been found to be beneficial to the user and more effective as a learning tool than traditional hypermedia, in terms of improved learning performance [4], and user-satisfaction [5].

Recently, another human factor, i.e. cognitive style, is being considered in AH systems. INSPIRE [20] and AES-CS [5] are two famous AH systems that take into account users' cognitive styles. In the former, users' cognitive styles are classified based on the model proposed by Honey and Mumford. This model, based on Kolb's theory of experiential learning, identifies learners as Activists, Pragmatists, Reflectors, or Theorists. In the latter, Witkin's Field Dependence is used to identify learners as Field Independent (FI), Intermediate (FM), or Field Dependent (FD). However, since these two systems adapt based on prior knowledge as well as cognitive style, reported benefits cannot necessarily be attributable to the adaptation to cognitive style. In this vein, this study was to examine whether the performance of an AH system can be enhanced by adapting to cognitive styles alone.

1.1 Cognitive Style

Cognitive styles refer to the way of how users process information. One of the most widely investigated cognitive styles with respect to hypermedia learning is field dependence. Field dependence refers to an individual's ability to perceive a local field as discrete from its surrounding field [8]. It is a single bi-polar dimension ranging from FD individuals at one extreme to FI individuals at the other.

Research has indicated differences in the way FD and FI individuals browse through hypermedia. For example, FD individuals tend to prefer a more restricted interface [9] and follow a linear route [10], whilst the converse is true for FI individuals. In addition, FD users have been found to prefer a breadth-first navigation path, where overviews of the topics are browsed first, whilst FI users prefer a depth-first path, browsing individual topics separately [12]. Further studies have highlighted differences regarding hypermedia structure and navigational aid preferences. FD users have been found to perform worse than FI users when there is no explicit structure within the interface [14], becoming confused and disorientated [15]. Furthermore, FD students have been shown to prefer using a map as a navigational aid [16], whilst FI users prefer an index [10]. Such studies are consistent with the conceptual differences between FD and FI individuals. Table 1 describes the relationships between the characters of FD and FI users and their navigation preferences.

Table 1. Field Independent vs. Field Dependence navigation preferences

Field Independent		Field Dependent	
Characteristic	Preference	Characteristic	Preference
Active approach	Prefer to use index to locate specific items	Passive approach	Rely on map to impose structure
Analytical tendency	Prefer depth-first paths	Global tendency	Prefer breadth-first paths
Internally Directed	Prefer non-linear and flexible navigation	Externally Directed	Prefer linear and restricted navigation

Based on Table 1, we have developed an adaptive hypermedia tutorial, which includes two types of interface: FI and FD interfaces (See Section 2.2). In addition, a normal interface that incorporated characteristics from these two interfaces was created. Comparing learning performance and user-perceptions of these three interfaces can help determine whether it is useful to consider cognitive styles in the development of AH systems. Therefore, this study aimed to examine this particular research question.

2. Methodology Design

2.1 Participants

64 participants took part in this experiment. All were second year Computer Science students at Brunel University and were each paid £5 for their participation and were

further motivated to take part in the experiment by being told that the tutorial may help them to learn the material from the course.

2.2 Materials

2.2.1 Web Tutorial

A Web tutorial was created to teach the students about computation and algorithms. This was split into two parts, one part of which was a standard tutorial with Normal Interface, the other adapted to suit either a FD or FI user. In order for some students to use the adaptive interface followed by the Normal interface, and others to use the adaptive interface followed by the Normal interface six half-tutorials were created (Normal, FD, FI for each half). The Normal interface was designed to be a richly linked hypermedia system to allow for non-linear learning. This meant the tutorial contained rich links within the text, as well as three navigation tools (a map, an index, and a menu) to aid the participants in their use of the tutorial.

Table 2. The differences of Field Independent and Field Dependent Interfaces

FI Interface	Adaptive Hypermedia	FD Interface
Depth-first path	Link Ordering	Breadth-first path
Rich Links	Link Hiding	Disabled Links
Alphabetical Index	Adaptive Layout	Hierarchical Map

Both FI (Figure 1) and FD (Figure 2) interfaces were developed on the basis of findings of previous hypermedia learning research that were summarized in Table 1. To achieve the particular aim of this study, these two interfaces only provided adaptivity and did not consider adaptability. In other words, the system automatically adapted the interface based on the users’ cognitive styles; the users were not able to customize the interface. As described in Table 2, three types of AH techniques were applied to develop these two interfaces, and their detailed functionalities are described below:

- *Link Ordering*: the system sorts a list of links according to users’ cognitive styles. In the FD interface, the links were sorted based on the Breadth-first path, which gave an overview of all of the material prior to introducing detail. In contrast, the FI interface took the Depth-first path, whereby each topic was presented exhaustively before the next topic, which was presented in the same way.
- *Link Disabling*: Due to the fact that FD users easily become disorientated and prefer to take a linear navigation strategy, the FD interface provided restricted navigation choices whereby links were disabled. On the other hand, the FI interfaces provided rich links, leaving freedom of navigation to the users.
- *Adaptive Layout*: Because FD and FI users process information in different ways, adaptive layout was applied to identify the relationships of the subject topics by providing different tools. The FD interface provided a hierarchical map, which could help the FD users to understand the content structure. Conversely, the FI interface used an alphabetical index to facilitate the location of specific information.

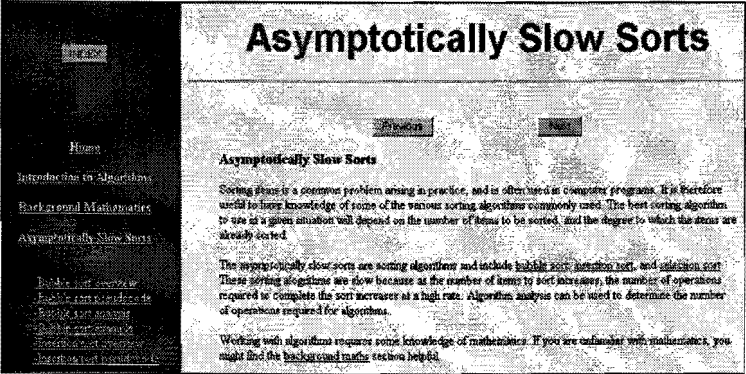


Fig. 1 Field Independent Interface

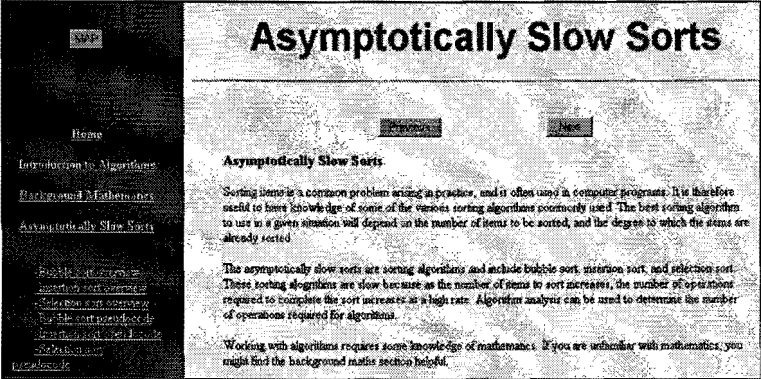


Fig. 2 Field Dependent Interface

2.2.2 Questionnaires

Two online questionnaires were created. The first of these (Questionnaire 1) asked for background information, such as age and gender, as well as information regarding the students' levels of prior knowledge of the subject domain (computation and algorithms knowledge). It also asked for their experience with, and their enjoyment using, computers, and the Web. Prior knowledge was measured on a 5-point scale to a series of questions related to how familiar with the subject the students were. The second questionnaire (Questionnaire 2) asked the students their perceptions of the Web tutorial. These included various questions regarding interface preference between the Normal and Adaptive interfaces, as well as questions regarding the user's ideal interface. This questionnaire, therefore, allowed for the analysis of a number of user perceptions of the interfaces and preferences between the two across a number of topics relevant to hypermedia learning.

2.2.3 Pre- and Post-Tests

Online pre- and post-tests were written to assess the participants' level of knowledge of the subject domain both before and after using the Web tutorial. Each test contained 20 multiple-choice questions on the subject, 10 of which were related to the first half of the tutorial, and 10 related to the second half of the tutorial. For each question there were five possible responses: four different answers and a "don't know" option. The questions were matched on the pre- and post-tests so that each question on the pre-test had a corresponding similar (but not the same) question on the post-test. Creating similar questions on the post-test was achieved by either re-writing the question or, where appropriate, by substituting different numbers into the questions.

2.2.4 Cognitive Style Analysis

A number of instruments have been developed to measure Field Dependence, including the Group Embedded Figures Test (GEFT) by Witkin et al. and the Cognitive Styles Analysis (CSA) by Riding. The main advantage of the CSA over the GEFT is that FD competence is positively measured rather than being inferred from poor FI capability. In addition, the CSA offers computerized administration and scoring. Therefore, the CSA was selected as the instrument in this study. In terms of the measures, Riding's recommendations are that scores below 1.03 denote FD individuals; scores of 1.36 and above denote FI individuals; students scoring between 1.03 and 1.35 are classed as Intermediate. In this study, categorizations were based on these recommendations.

2.3 Design

In order to determine whether or not the adaptive interface was better than the normal interface a within-subjects design was used. This meant that each student used both the normal interface and an adaptive interface. To avoid a learning effect, each of these interfaces covered different topics within the tutorial. Since the interfaces were on different topics within the tutorial it was necessary to create both adaptive and normal interfaces for each of the two half-tutorials, so that half of the students used the normal interface for the first half of the tutorial and the adaptive interface for the second half of the tutorial. Similarly the other half of the students used the adaptive interface for the first half of the tutorial and the normal interface for the second half. This was necessary to avoid any effects of interface preference being related to the content of the interface rather than its presentation.

Finally, in order to show that any effects of interface preference were related to matching with the user's cognitive style rather than just a preference for any adaptive interface, users were randomly matched or mismatched to their cognitive styles: approximately half of the participants used the adaptive interface that was suited to their level of field dependence, whilst the other half used the adaptive interface that was the opposite to the one they were suited to. This meant that for any student there were four possible experimental conditions: FD interface followed by Normal interface, FI/Normal, Normal/FD, Normal/FI.

2.4 Procedure

The experiment was carried out over a number of sessions in December 2003. The students took part in one session only. Each session contained a small group of students each working individually. The experiment began by the student taking the CSA to determine their level of field dependence. This was used to automatically provide adaptation of the tutorial interface to suit the user's level of field dependence. Students were randomly assigned to an interface that was either matched with their cognitive style or mismatched with it. After the CSA the students completed Questionnaire 1. This was followed by the Pre-test. This was timed allowing the students a maximum of 15 minutes to complete. Students could submit their test before the 15 minutes was up, but once the time was up the system automatically submitted the test and proceeded to the next section. The Pre-test was followed by using the first interface of the tutorial for 25 minutes, and then the second interface for 25 minutes. This was then followed by the Post-test, again with a 15 minute time limit, before finally with Questionnaire 2.

2.5 Data Analyses

The independent variable was the user's level of field dependence as measured by the CSA. The dependent variables were the responses to the various questions about the tutorial from Questionnaire 2, as well as learning performance based on the tests. All questionnaire responses, where appropriate, were scored as 5 for "strongly agree", through to 1 for "strongly disagree". A "gain score" was calculated as the post-test score minus the pre-test score.

SPSS for windows was used for the analysis of the data. Pre- and post-test scores were given as marks out of 20. A significance level of 0.05 was adopted. Chi-square tests were used to analyze interface preference in the matched and mismatched conditions, since this data was in the form of frequencies. Pearson's correlations were used to analyze the relationship between field dependence and questionnaire responses, where field dependence was measured on the continuous score as given by the CSA, as opposed to the discrete categories of FD and FI.

3 Results and Discussion

3.1 Interface Preference

Analysis of participants' interface preference indicated that there was no significant preference between the Normal interface and the Adapted interface for the participants who were matched with their cognitive style. However, those who were mismatched to their cognitive style were significantly more likely to prefer the Normal interface over the Adapted interface (chi-square = 5.26, $df = 1$, $p < 0.05$). Figure 3 highlights this finding. This finding suggests that there may be an important interaction between field dependence and interface preference. However, whilst the users were significantly more likely to prefer the Normal interface over the Adapted

interface when they were mismatched with their cognitive style, there was no significant preference for the Adapted interface when the users were matched with their cognitive style (with approximately half preferring the Adapted interface and half the Normal interface).

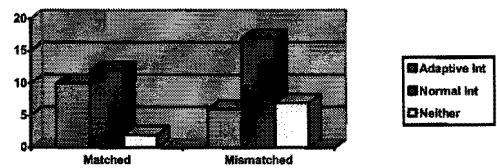


Figure 3. Preferences in matched and mismatched conditions

This suggests that whilst a wrongly adapted interface may cause problems for some users, appropriately adapted interfaces may be no more effective than a well-designed interface for all users. This is consistent with other studies adopting a matched/mismatched design (e.g., [12]), which have shown mismatched participants to experience more difficulties than matched participants. It is possible that the Normal interface in this study contained positive aspects for both FD and FI users. For example, the Normal interface provided links within the text that would be suitable for a FI user, whilst also having next/previous buttons to provide direct guidance for FD users. Also, the Normal interface contained both a map and an index. Supporting this conclusion is the fact that 75% of all the participants (including 58 % of FD participants and 77% of FI participants) preferred having a selection of navigation tools. This finding contrasts with previous research indicating that FIs prefer an index and FDs a map (e.g., [10][16]). Whilst it is possible that FDs do prefer a map, and FIs an index, from this study it seems that overall users prefer a selection of navigation tools.

This study, thus, poses the question of whether it is possible to create a single interface that can be suitable for both FD and FI users. Whilst it is possible that the adapted interfaces in this study could be further improved to make them better than the Normal interface, it is important for further studies to determine whether adapted interfaces can be created that are genuinely beneficial above a single interface used by all. With the findings of this study in mind, it is possibly more beneficial for hypermedia system designers to concern themselves with an interface that is easy (and not too restrictive) to use for all users, regardless of their level of field dependence. Trying to create distinct interfaces for different levels of field dependence may do more harm than good. Since field dependence is measured on a continuous scale and is only superficially grouped into distinct categories, it is difficult to decide categorically the preferences of any given user. Whilst some users may prefer an interface that is consistent with the literature regarding their level of field dependence, others may not be. For example, a user at one extreme of the scale may prefer a different interface to a user in the same category, but with a less extreme score. A more suitable interface would be one that was neutral and could support all users, whilst allowing the user to specify any particular changes that they would like

(as with the AES-CS [5]). Such an interface would hopefully provide the user with a suitable interface, whilst alleviating any particular difficulties they may have.

Despite the finding that mismatched users preferred the Normal interface, Chi-squared tests carried out between FD/intermediate/FI and six other questions referring to aspects of interface preference showed just one significant finding. FI participants found it easier to get lost using the adaptive interface than the Normal interface (chi-square = 4.8, $df = 1$, $p < 0.05$). However, since significance was not even approached for FD participants or intermediates, nor for the similar questionnaire responses regarding interface navigation, it seems likely that this result is anomalous. Furthermore, analysis of learning performance as measured by the post-test score minus the pre-test score showed no significant difference on learning performance using the adaptive interface between those who were matched and those who were mismatched. In fact, the results indicated that those who were mismatched performed marginally better (mismatched mean gain score = 1.1, matched mean gain score = 0.96). In this respect this experiment is inconsistent with the majority of reported studies (e.g., [16]). However, it is consistent with those studies that found no significant differences in learning performance (e.g., [17]).

3.2 Ideal Interface Perceptions

Pearson's correlations carried out between field dependence score and six questions referring to what the user thought the ideal interface should contain found one significant correlation. Field dependence score was correlated with the statement 'how important do you think the following features are to a tutorial: Providing an example of an algorithm first, before giving more detail' ($r = .267$; $p < .05$). This indicated that FD users found providing an example first more important than did the FI users. This result is consistent with previous research [12], and justifies the FD interface directing the user with an example before giving more detail.

However, it is perhaps surprising that none of the other statements showed any significant correlations, since these were also considered to be characteristic of one or other of the cognitive styles. This suggests that the differences between FD and FI users (as measured by the CSA) in terms of hypermedia preferences may not be as strong as previously believed, at least in terms of subjective preferences. Previous research has suggested that FD users prefer to follow a linear route through hypermedia, whilst FI users prefer to be more flexible (e.g., [9]). Yet, no such correlation was found in this study. Such results would have important implications for designing AH systems to adapt to cognitive style. Since differences may not be clear cut, adaptation to an interface that is too rigidly 'FD' or 'FI' may not be beneficial, and may not suit the preferences of the individual user. In particular, since only one significant difference was found between FD and FI users in relation to ideal interface design, it is important to determine whether the needs of FD and FI users are as clear-cut as are claimed.

4. Conclusion

The results from this study suggest that it is possible that adapting to an interface that is too specific may restrict users, who may not necessarily prefer all aspects of the interface that are considered to be useful for a user with that cognitive style. In this study the Normal interface was less restrictive and may have suited users of both cognitive styles. It appears that the Normal interface incorporated enough freedom of navigation to suit those who preferred to navigate freely, whilst also providing a suggested route for those who needed structure. It also provided a range of navigation tools that was found to be preferable by the majority of the users to having just one.

Since this experiment was restricted to the study of field dependence as measured by the CSA, future research should also strive to determine whether the findings from this study regarding adaptation to field dependence apply to other cognitive styles and other cognitive style assessment tools. Furthermore, this study was limited in that it provided adaptation to field dependence and field independence in a way considered appropriate for such individuals based on interpretations of previous research into field dependence and hypermedia learning. Since some findings from this study differ from this interpretation (for example, users preferring a selection of navigation tools as opposed to just one), future studies might consider revising the interpretation used here concerning ideal interfaces. Future research could therefore re-interpret what an ideal interface might be for FD and FI users, and determine whether different interfaces are needed, or whether one could satisfy all users regardless of their level of field dependence.

Acknowledgements

The authors are grateful for the support of the EPSRC (Grant number GR/R57737/01) for funding this project.

References

- [1] Wu, H., De Bra, P., Aerts, A., Houben, G.J.: Adaptation Control in AH Systems. Proceedings of the AH2000 Conference (2000) 250-259
- [2] Brusilovsky, P.: Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology* 34(4) (2003) 487-497
- [3] Schwarz, E., Brusilovsky, P., Weber, G.: World wide intelligent textbooks. Paper presented at the World Conference on Educational Telecommunications, June 1-22, Boston, MA (1996)
- [4] Boyle, C., Encarnacion, A.O.: Metadoc: an adaptive hypertext reading system. *User Modeling and User-Adapted Interaction* 6 (1994) 1-19
- [5] Triantafillou, E., Pomportsis, A., Demetradis, S., Georgiadou, E.: The value of adaptivity based on cognitive style: an empirical study. *British Journal of Educational Technology* 35(1) (2004) 95-106
- [6] Brusilovsky, P.: Adaptive Educational Systems on the World-Wide-Web: A Review of Available Technologies. In: Proceedings of Workshop "WWW-Based

- Tutoring" at 4th International Conference on Intelligent Tutoring Systems (ITS'98), San Antonio, TX (1998)
- [7] Brusilovsky, P., Pesin, L.: Adaptive navigation support in educational hypermedia: An evaluation of the ISIS-Tutor. *Journal of Computing and Information Technology* **6**(1) (1998) 27-38
 - [8] Witkin, H. A., Moore, C.A., Goodenough, D.R., Cox, P.W.: Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research* **47**(1) (1977) 1-64
 - [9] Dufresne, A., Turcotte, S.: Cognitive style and its implications for navigation strategies. In Boulay, B., Mizoguchi, R. (eds.): *Artificial intelligence in education knowledge and media learning system*. Amsterdam IOS Press, Kobe, Japan (1997) 287-293
 - [10] Liu, M., Reed, W.M.: The effect of hypermedia assisted instruction on second-language learning through a semantic-network-based approach. *Journal of Educational Computing Research* **12**(2) (1995) 159-175
 - [11] Chen, S. Y., Ford, N.: Modelling User Navigation Behaviours in a Hypermedia-Based Learning System: An Individual Differences Approach. *International Journal of Knowledge Organization* **25**(3) (1998) 67-78
 - [12] Ford, N., Chen, S.Y.: Matching/mismatching revisited: An empirical study of learning and teaching styles. *British Journal of Educational Technology* **32**(1) (2001) 5-22
 - [13] Chang, C.T.: A study of hypertext document structure and individual differences: Effects on learning performance. PhD. Dissertation. University of Illinois at Urbana-Champaign (1995)
 - [14] Palmquist, R.A., Kim, K.-S.: Cognitive style and on-line database search experience as predictors of Web search performance. *Journal of the American Society for Information Science* **51**(6) (2000) 558-566
 - [15] Wang, P., Hawk, W.B., Tenopir, C.: Users' interaction with World Wide Web resources: An exploratory study using a holistic approach. *Information Processing & Management* **36** (2000) 229-251
 - [16] Ford, N., Chen, S.Y.: Individual differences, hypermedia navigation and learning: An empirical study. *Journal of Educational Multimedia and Hypermedia* **9**(4) (2000) 281-312
 - [17] Fitzgerald, G.E. Expert and novice performance in a hypermedia learning environment for classroom observation skills: The impact of prior teaching experience. *Journal of Educational Computing Research* **19**(2) (1998) 115-131
 - [18] Boyce, K.E.: Delivering continuing professional education at a distance: The correlation of field dependence/independence and learning using the World Wide Web (Field Dependence, Distance Education). PhD Dissertation. The University of Oklahoma (1999)
 - [19] Umar, I.N.: A study of the effects of cognitive styles and learning strategies among Malaysian pre-college students in a hypermedia environment. Ed.D Dissertation. University of Pittsburgh (1999)
 - [20] Papanikolaou, K. A., Grigoriadou, M., Magoulas, G. D. & Kornilakis, H. (2002) Towards new forms of knowledge communication: the adaptive dimension of a web-based learning environment, *Computers & Education*, 39(4), 333-360

WORKSHOP 3:
Empirical Evaluation of Adaptive Systems

Workshop Co-Chairs:

Stephan Weibelzahl
Alexandros Paramythis

Preface

Only few publications on user modeling systems report significant evaluation results. One of the reasons that have been identified to be responsible for this lack is that both methods and criteria for assessing the "success" of adaptation are not clear. However, user modeling systems are designed for human-computer interaction and thus, empirical approaches for such assessment are absolutely necessary in order to justify the efforts required to render systems adaptive.

This is the third workshop on empirical evaluation of adaptive systems. The workshop's guiding perspective is that adequate methods and reliable criteria are prerequisites towards increasing the quantity and quality of evaluation studies on adaptive systems. The workshop aims to contribute to the exploration and discussion of suitable methods and criteria in various domains with differing user modeling and adaptation techniques. It further aims to encourage researchers to perform evaluation studies with their own adaptive hypermedia systems.

The workshop, continuing in the steps of its predecessor, will focus on the following questions:

- Which of the existing empirical criteria and methods are appropriate for the evaluation of adaptive hypermedia systems? What new criteria need to be introduced to specifically cater for the presence of adaptation in the evaluated systems? What empirical methods are appropriate (or, how do existing methods need to be modified, so as to be suitable for) assessment against the new sets of criteria?
- Can metrics be developed to facilitate the comparison between different (versions of) adaptive systems, or between adaptive and non-adaptive systems?
- How can we foster an increase in the volume and quality of empirical evaluations of adaptive systems? What are the most common pitfalls that can be identified in previous studies?

August, 2004

Stephan Weibelzahl and Alexandros Paramythis

Program Committee

Peter Brusilovsky, *University of Pittsburgh, USA*

Judith Masthoff, *University of Brighton, UK*

Alexandros Paramythis, *Johannes Kepler University Linz, Austria*

Gerhard Weber, *University of Education Freiburg, Germany*

Stephan Weibelzahl, *Fraunhofer IESE, Germany*

Proposed Evaluation Framework for Adaptive Hypermedia Systems

Arpita Gupta, P. S. Grover

Department of Computer Science, University of Delhi,
Delhi – 110 007, India
arpita_aggarwal@hotmail.com, psgdelhi@yahoo.com

Abstract. Although a number of frameworks exist for the evaluation of Adaptive Hypermedia Systems (AHS), recently suggested layered frameworks have proved useful in identifying the exact cause of the adaptation failure or any other error in the system. This paper presents an evaluation framework for AHS for internet which is an extension of the layered frameworks and adds new dimensions to them. It treats evaluation as an integral part of development process of AHS and also evaluates the successful access of AHS on the internet. The framework has four dimensions which are orthogonal to each other – Environment – the environment in which AHS is accessed, Adaptation – the type of adaptation used, Development Process – software engineering life cycle steps used for developing AHS and the Evaluation Modules – the layers of AHS which are evaluated in context of other dimensions.

1 Introduction

Adaptive Hypermedia Systems (AHS) are designed and built with the intention of providing tailor made information to individual users according to their preferences, goals and knowledge. With the advent of internet as a common source of information, they have found a platform to reach heterogeneous groups of users using different devices for assessing AHS. With this increases the challenge of catering to a wide variety of users in differing environments and also the added responsibility of working without making mistakes since a single mistake can make the user lose trust in the system – maybe forever. Therefore, evaluating the AHS is of utmost importance. Moreover, it is equally important to have a correct method of evaluation since an incorrect method can lead to wrong conclusions [3].

Earlier evaluation studies compared adaptive versions of the system with the non-adaptive versions [2, 4]. A major criticism of this approach was that the non-adaptive versions – usually implemented using adaptive version with their adaptivity switched off – were not “optimal” [11].

Recently some layered evaluation frameworks were suggested which do not treat evaluation as a “monolithic” process but instead divide it into layers [3, 20, 23]. This

approach helps in identifying the exact cause of the adaptation failure or any other error. These evaluation frameworks basically differ in layer granularity and do not take “extensibility or maintainability” of the AHS into consideration [10]. Another evaluation framework – Extended Abstract Categorization Map (E-ACM) [21] has been suggested to guide adaptation evaluation and design. Most of them do not take the development process into consideration resulting in detection of errors and weaknesses in the system which can prove to be expensive to correct at a later stage. Moreover issues like maintenance of the system, the environment in which they will be used – location and devices accessing AHS, etc., have not been addressed in the current frameworks.

There can be a number of factors which affect the evaluation process of an AHS. Internet provides the opportunity of accessing AHS using a variety of devices like desktops, mobiles, PDAs, in any location of the world, to the users having diverse skills, capabilities and knowledge. The AHS itself can belong to any application domain having some specific characteristics. All these elements form the environment of the AHS. The adaptation in the AHS can be static or dynamic depending on the time and process of adaptation.

Similar to software engineering, AHS also involves the analysis, design, implementation and maintenance phases of development process. During these phases, they should be evaluated for the validity of input acquisition, correctness of inferences drawn from these inputs, correctness of various models created by the AHS, the adaptation decisions taken based on these models and their final presentation to the user.

We propose an evaluation framework treating the evaluation as an integral part of the development process of AHS and taking the accessing environment and the type of adaptation provided by AHS into consideration while evaluating individual modules of AHS. The framework consists of four orthogonal dimensions: Environment, Adaptation, Development process and the Evaluation modules. Next Section describes the framework.

2 Proposed Evaluation Framework

The proposed framework is an extension of layered evaluation frameworks where the layers need to be evaluated in context with other dimensions. The benefits of the framework are: (i) it allows a structured, layered view to better understand the various aspects of AHS (ii) it can be used as a conceptual framework for evaluating existing approaches for AHS (iii) it may be used during the development of next generation AHS using software engineering steps.

The framework consists of 4 dimensions - Environment, Adaptation, Development Process, and Evaluation Modules. These dimensions are orthogonal to each other i.e. all the evaluation modules should address all the components of environment and adaptation during each phase of development process.

Figure 1 shows our proposed evaluation framework and the following sections describe these dimensions and their components.

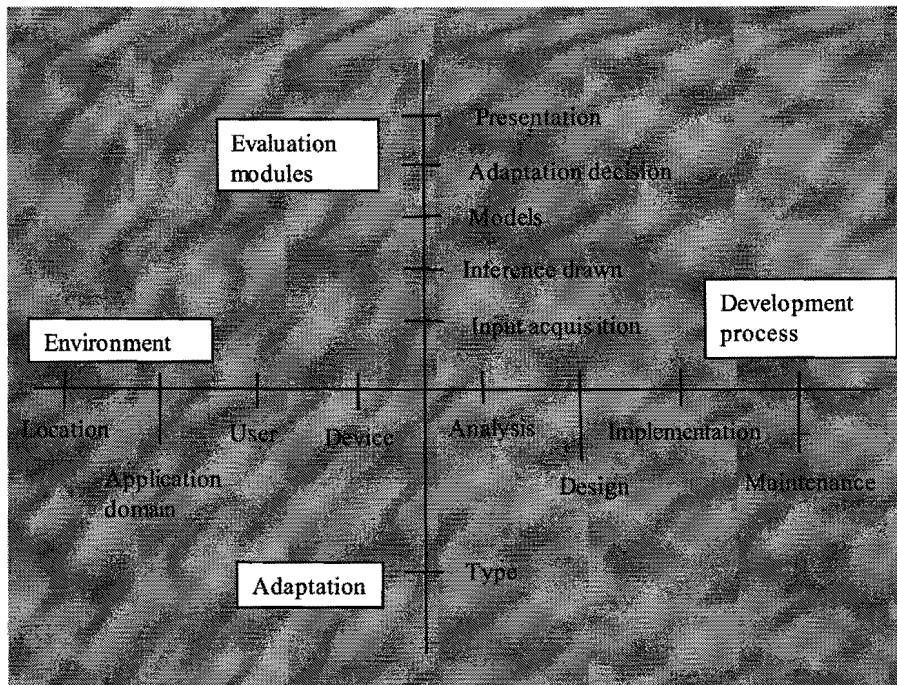


Figure 1. Proposed Evaluation Framework for AHS

2.1 Environment

The first dimension – environment – is the set of conditions to which AHS has to adapt itself i.e. they are the circumstances of consumption for AHS. There can be innumerable variables affecting the environment; we group them under the following components:

2.1.1 Device: Advent of web-capable appliances with limited abilities such as PDAs and mobile telephones along with desktops – have made one-size-fit-all paradigm obsolete since the range of hardware and software used at client side of web-based systems is extremely wide. AHS need to be evaluated for the correct acquisition of the device characteristics and smooth running with hardware features like display sizes, local storage size, method of input, processing speed and software features like browser versions, available plug-ins, Java and JavaScript etc. Kobsa et al [16] have discussed acquisition of such data by the AHS.

2.1.2 User: The personal characteristics of user such as demographic data, user knowledge, user skills and capabilities, user interests and preferences, his goals and plans have been used by many AHS for adaptation [2, 22]. Along with these features,

AHS should be evaluated for adaptation according to all users including disabled and elderly users [15].

2.1.3 Application Domain: AHS can be developed for a wide range of applications which differ in characteristics. Brusilovsky [5] have identified several application domains and existing AHS for them. The application specific characteristics constitute critical parameters while evaluating AHS and should be evaluated along with general characteristics of AHS since the traditional concerns change with the application domain.

2.1.4 Location: Information about the geographical location of the accessing device can be used to filter and adapt or recommend the content of AHS and should be evaluated for the correct delivery of the same. Kobsa et al [16] have given methods for acquisition of information about the location and their use in adaptation procedure.

2.2 Adaptation

The second dimension of the framework is adaptation which can be of two types: **static** adaptation and **dynamic** adaptation, depending upon the time and process of adaptation. Static adaptations are specified by the author at the design time or determined once only at the startup of the application. Fink et al [9] used static adaptation in AVANTI project. Dynamic adaptation occurs during runtime depending on various factors like inputs given by the users during use, changes in user model, adaptation decision taken by AHS etc. Kappel et al [13] have distinguished three options for dynamic adaptation – immediate dynamic adaptation i.e. adaptation done as soon as context changes, deferred dynamic adaptation i.e. adaptation done only after the user has requested the page which is subject to adaptation, and periodic adaptation i.e. adaptation is done periodically. Example of system having dynamic adaptation is AHA [8].

2.3 Evaluation Modules

The third dimension of evaluation framework consists of evaluation modules which need to be considered for evaluation of AHS. These modules have been suggested in the layered frameworks [3, 20, 23]. Our perspective is to evaluate them with respect to other dimensions of the framework.

2.3.1 Input Acquisition: Inputs are required from the environment as well as from the user. These can be taken manually (i.e. user feeds them), automatically (i.e. system takes the input itself e.g. Type of device, its screen size, location etc) or semi automatically (i.e. combination of both – some input through user, some automatically) [13].

The inputs taken by the system – either manually or automatically might not carry any semantic information, but they need to be evaluated for the reliability, accuracy, precision, latency, sampling rate, so that the inferences drawn from them are valuable.

This needs to be done at all stages of development process – analysis, design, implementation and maintenance phases, for both static and dynamic adaptations, for all intended devices, users, locations and application domains.

2.3.2 Inferences Drawn: Previous layer was involved with the data collection, this layer gives “meaning” or “semantics” to it i.e. it draws inferences from it. Evaluators need to check if these inferences or the conclusions drawn by the system concerning the user-computer interaction are correct since it is not necessary that there will be a direct – one to one mapping between raw data and their semantically meaningful counterparts.

Moreover, inputs given by various users of different devices or application domain might need different interpretations. Evaluators need to check if all such interpretations have been analyzed, designed and implemented in the AHS for both static and dynamic adaptations.

2.3.3 Models: For achieving the required adaptations, various models are created by the system. Benyon and Murray [1] specified three models – user model, domain model, interaction model. Nora Koch [18] has described four models for carrying out the adaptation – user model, navigation model, presentation model, and adaptation model.

These models are based on the inferences drawn in the previous stage and are suppose to imitate the real world. They need to be evaluated for validity i.e. correct representation of the entity being modeled, comprehensiveness of model, redundancy of model, precision of the model, sensitivity of the modeling process [20].

2.3.4 Adaptation Decision: Given a set of properties in the user model, sometimes there can be more than one adaptation possible. In this module, evaluation of the most “optimal” adaptation is done using criteria like necessity of adaptation, appropriateness of adaptation, acceptance of adaptation [20]. Careful evaluation is needed to ascertain that increase in adaptation is not resulting in decreased usability [5].

2.3.5 Presentation: This module involves the human-computer interaction and needs to be evaluated for criteria like completeness of the presentation, coherence of presentation, timeliness of adaptation, user control over adaptation [20].

2.4 Development Process

The fourth dimension of the framework is the development process comprising of phases of software life cycle i.e. analysis, design, implementation and maintenance. Benyon and Murray [1] gave a star approach to interactive system development taking evaluation as central element and system analysis, specification of user requirements, design, prototype and implementation as the peripheral elements.

During each phase of this dimension, evaluation of individual elements of environment, adaptation and evaluation modules is done with respect to each other.

2.4.1 Analysis: This phase involves gathering information about the problems of current system, and/or identifying the requirements and constraints of the system to be developed. Main components of this phase are:

- *Functional analysis:* establishes the main functions that the system is expected to perform and how it is to perform.
- *Environment analysis:* This analyzes the environment in which the system is expected to be accessed – including the physical aspects like location, device and other aspects like application domain, type of user.
- *User and task analysis:* This determines the scope of cognitive characteristics like user's preferences, goals, knowledge and other attributes required in user model e.g. search strategy required, assumed mental model etc.
- *Interface analysis:* It identifies features like effectiveness, learnability, flexibility and attitude required of the system.
- *Data analysis:* This involves the analysis of input acquisition to identifying the data to be stored and manipulated by the system, and to understand and represent the meaning and structure of data in the AHS.
- *Analysis of Models:* This involves the analysis of various models maintained by the system such as user model, domain model, navigation model, adaptation model.

To evaluate this phase, checklists can be prepared for different types of analysis mentioned above, corresponding to various components of different dimensions of the evaluation framework. For example, a checklist is prepared for the desktop computer's functional requirements of static adaptation for input acquisition; another checklist for the PDAs for the same specifications is prepared.

2.4.2 Design: Design phase defines the overall system organization by transforming the functions and tasks defined during analysis phase into software components and their externally visible properties of those components and their relationships. It is recommended to design the adaptive parts of the system in parallel with the whole system so as to have a successful adaptation [11]. Some of the components for this phase are:

- *Architectural Design:* Many architectural designs have been suggested for the adaptive systems [1, 12, 19]. The modular architecture model presented in figure 2 is especially designed for our evaluation purpose.

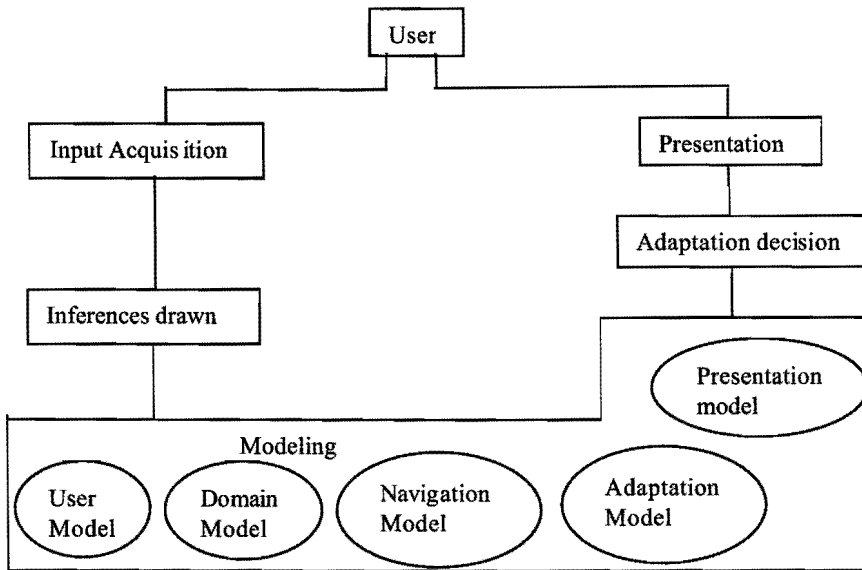


Figure 2. Architecture model for evaluation of adaptive hypermedia system

In this model, “Input acquisition” module acquires input by the user through keyboard or mouse or takes it from the environmental factors such as the device and location accessing the AHS. This input is verified and then passed on to the next module “Inferences drawn” for drawing logical inferences and conclusions which are then stored in some dynamic models created by the system such as user model. The “modeling” module consists of various models – both static and dynamic such as user model, domain model, navigation model, presentation model and adaptation model. Different AHS can have one or more such models which are used at various levels of adaptation. Depending on some these models, “adaptation decision” module takes the decision so as to choose the best adaptation techniques out of the available ones in the system. Then the “presentation” module presents the final content and links to the user.

Evaluation of architectural design is important because it defines constraints on implementation and maintenance, making it easier to reason about and manage changes, and defines system’s quality attributes.

- *Content Design:* Content forms one of the most important aspect of the AHS since it is for the content that the AHS exists. Therefore, it should be verified for accuracy and precision. Moreover, there is a need to identify and evalu-

ate the content which would be visible to different users (according to their individual user model), on different devices and location.

- *Navigation Design:* Since a number of navigational techniques are present such as direct guidance, link hiding etc [5] which can be used with various navigation aids like icons, graphics, there is a need to ascertain that the chosen one is appropriate for content and consistent with the heuristics leading to high quality interface designs and also ensure that aids like site maps and table of contents are designed into the system.

Evaluation of navigational design is needed to ensure the maintenance of coherence of the overall system as the user moves across application since it is related to AHS's underlying communicative performance.

- *Interface Design:* Since user interface is the "first impression" of the system, a poorly designed interface might disappoint user regardless of the value of content, sophistication of architectural design and navigation techniques used. Therefore, careful evaluation is necessary for well structuredness and ergonomically sound interface designs.

Design phase can be evaluated by using metrics such as structural complexity metrics, navigational metrics, usability metrics etc. Moreover, a good architecture exhibits low coupling and high cohesion in terms of some decomposition of functionality [14].

2.4.3 Implementation: During the actual implementation of the system, evaluation of each module of AHS should be carried out individually and then in integration with the other modules for the successful adaptation at various levels – both static and dynamic, for different users, on different devices. Some metrics like behavioral complexity, reliability metrics, precision, software size and length metrics help in evaluation of the system as a whole. Moreover, evaluation of tools can be done in relation to the activities in which they are used to indicate their availability for each kind of activity and how the tool supports it.

2.4.4 Maintenance: AHS might require updation at any moment; therefore, during the design phase care should be taken to design hyperspace in modular way so that change in structure can be made by changing the relations among these modules. Automatic link generation should be preferred over static links to preserve the link coherence. Maintenance of static links is very complex as any change in a node position in hyperspace necessitates it to revise all the documents that include links to this node, in order to update them.

Content maintenance can be made easy by storing contents apart from the concept structure or the navigational options since contents are external and easily updateable. Finally, database should be used to store information about different system components to facilitate the management and maintenance of these components and guarantee data consistency [7]. Checklists can be prepared for these and metrics can be used

to measure the ease of maintenance such as complexity metrics, reuse metrics or expandability metrics.

Table 1 gives an example to illustrate the way in which to use the framework. While developing an adaptive tutoring system for internet, during the implementation phase, the precision of the content is measured for the presentation module and the values are filled in the table 1. The values are filled by using various metrics appropriate for different stages and can be in any unit of measurement. For the complete evaluation, many such tables would be required for various phases.

Table 1. An Example to show how to use the framework dimensions (values are only for demonstration purpose and are not exact).

Precision of content during implementation phase of presentation module		Adaptation		
			Static Ad-aptation	Dynamic Adaptation
Environment	Device	PDA	85%	80%
		Mobile	80%	80%
		Desktop	95%	96%
	User	Novice	70%	76%
		Average	90%	90%
		Expert	80%	85%
	Location	.		
		.		
		.		
	Application Domain	.		
		.		
		.		

3 Conclusion

Our proposed evaluation framework integrates the AHS development process, the accessing environment, the different types and levels of adaptations involved in AHS and the evaluation modules of layered frameworks. Several factors that impact the AHS evaluation can be organized around these framework perspectives.

The framework is a mechanism to gain insight into and an understanding of AHS for internet. It can be used for summational evaluations once the AHS has been completed by replacing the “development process” with “initial goals and achieved goals” and checking with the rest of the three dimensions. It can also be used for formative evaluations during the development of the system by establishing goals for each phase and then compare the actual results.

The dimensions and their elements suggested in the framework have been addressed more or less globally. Subfactors can be established for each element which can be evaluated objectively or subjectively.

We are in the process of developing and evaluating an adaptive tutoring system with the authoring tool AHA using this framework where analysis phase has checklists and set of goals prepared according to the requirements. Metrics largely are being used during design, implementation and maintenance phases for the purpose of evaluation. The results of this study will be reported later on.

4 References

1. Benyon D., Murray D.: Adaptive Systems: from intelligent tutoring to autonomous agents. *Knowledge-Based Systems*, 6(4), 197-219 (1993)
2. Boyle C., Encarnacion A. O.: Metadoc: An Adaptive Hypertext Reading System. In P. Brusilovsky et al. (Eds.), *Adaptive Hypertext and Hypermedia*, ©1998 Kluwer Academic Publishers, pg 71-89, (1994)
3. Brusilovsky P., Karagiannidis C., Sampson D.: The Benefits of Layered Evaluation of Adaptive Applications and Services. In Weibelzahl S., Chin D. N., and Weber G. (eds), *Empirical evaluation of Adaptive systems*, Proceedings of workshop at the eighth international conference on user modeling, UM2001, pg 1-8, Freiburg
4. Brusilovsky P., Eklund J.: A Study of User Model Based Link Annotation in Educational Hypermedia. *Journal of Universal Computer Science*, vol. 4, no. 4 (1998), pg 429-448, (1998)
5. Brusilovsky P.: Methods and Techniques of Adaptive Hypermedia, P. Brusilovsky et al (eds.), *Adaptive Hypertext and Hypermedia*, 1-43, Kluwer Academic Publishers, 1998, printed in the Netherlands
6. Brusilovsky P.: Efficient Techniques for Adaptive Hypermedia, In C. Nicholas and J. Mayfield (eds): *Intelligent hypertext: Advanced techniques for the world wide web*. LNCS, 1326, Berlin:Springer-Verlag, 12-30
7. Carro, R.M.: Adaptive Hypermedia in Education: New Considerations and Trends. *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics* (Orlando, Florida), Vol. 2, ISBN: 980-07-8150-1, 452-458, (2002)
8. De Bra, P., Smits D., Stash N.: AHA! The next generation. *ACM Conference on Hypertext and Hypermedia*, May 2002
9. Fink, J., Kobsa, A., Nill, A.: Adaptable and adaptive information provision for all users, including disabled and elderly people. In *The New Review of Hypermedia and Multimedia*, 4, 163-188
10. Gupta A., Grover P. S.: Comparison of Evaluation Frameworks for Adaptive Hypermedia. To be published in *Proc. of 2nd International Conference on Quality, Reliability and Information Technology (ICQRIT Dec 2003)*, New Delhi
11. Höök K.: Steps to take before intelligent user interfaces become real. *Interacting with Computers*, 12, pg 409-426, (2000)
12. Jameson, A.: *Systems That Adapt to Their Users: An Integrative Perspective*. Saarbrücken: Saarland University, (2001)

13. Kappel, B. Pröll, W. Retschitzegger, W. Schwinger: Customisation for Ubiquitous Web Applications - A Comparison of Approaches. *Int. Journal of Web Engineering and Technology (IJWET)*, Volume 1, No. 1, 2003, pp. 79-111, [ISSN 1476-1289]
14. Kazman R., Clements P., Bass, L.: *Software architecture in Practice*. Addison-Wesley, 1998, pg 218
15. Kobsa, A., Stephanidis, C. : Adaptable and Adaptive Information Access for All Users, Including Disabled and Elderly People. *Proceedings of 2nd Workshop on Adaptive Hypertext and Hypermedia, HYPERTEXT'98*, Pittsburg, USA, June 20-24 (1998)
16. Kobsa A., Koenemann J., Pohl W.: Personalised Hypermedia Presentation Techniques for Improving Online Customer Relationships. *The Knowledge Engineering Review*, Vol. 16:2, 111-155, (2001), Cambridge University Press
17. Mendes E., Hall W., Harrison R.: Applying Metrics to the evaluation of Educational Hypermedia Applications, *Journal of Universal Computer Science*, vol. 4, no. 4 (1998), pg 382-403
18. Nora Koch, PhD Thesis : Software Engineering for Adaptive Hypermedia Systems: Reference Model, Modeling Techniques and Development Process. Ludwig-Maximilians-University of Munich, Germany, December 2000
19. Oppermann, R.: Adaptively supported adaptability. *International Journal of Human Computer Studies*, 40(3), 455-472, (1994)
20. Paramythis A., Totter A., Stephanidis C.: A Modular Approach to the Evaluation of Adaptive User Interfaces, in Weibelzahl, S. Chin, D.N., Weber, G. (eds), *Empirical Evaluation of Adaptive Systems*, *Proceedings of workshop at the eighth International Conference on User Modeling, UM2001*, pg 9-24, Freiburg, (2001)
21. Tobar, C. M.: Yet Another Evaluation Framework. In: Weibelzahl, S. and Paramythis, A. (eds.). *Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems*, held at the 9th International Conference on User Modeling UM2003, Pittsburgh, (2003), pp. 15-24
22. Weber, G, Specht, M.: User modeling and adaptive navigation support in WWW-based tutoring systems.. In a Jameson, C Paris and C Tasso (eds) *User Modeling: Proceedings of the sixth International Conference* Springer-Verlag, (1997), 289-300
23. Weibelzahl S., Lippitsch S., Weber G. (2002): Advantages, Opportunities, and Limits of Empirical Evaluations: Evaluating Adaptive Systems, *Künstliche Intelligenz*, 3/02, 17-20

The First Click is the Deepest: Assessing Information Scent Predictions for a Personalized Search Engine

Karen Church, Mark T. Keane & Barry Smyth

Adaptive Information Cluster, Department of Computer Science,
University College Dublin, Belfield, Dublin 4, Ireland
{karen.church, mark.keane, barry.smyth}@ucd.ie

Abstract. “First-click behavior” describes one of the most commonly occurring tasks on the Web, where a user submits a query to a search engine, examines a list of results and chooses a link to follow. Even though this task is carried out a billion times a day, our understanding of the factors influencing this behavior is poorly developed. In this paper, we empirically evaluate information scent predictions for first-click behavior in the use of a personalized search engine, called I-SPY. Our experiments show that the predictive accuracy of current information foraging approaches is not good. To conclude, we advance a framework designed to understand first-click behavior and guide future research.

1 Introduction

Almost every time someone opens a web browser they carry out a very simple behavioral sequence leading to their first click to some distant website. This sequence involves the user submitting a query to some search engine, scanning a list of returned results and choosing to click on a selected link. Though this, apparently simple, “first-click behavior” is incredibly commonplace, it is still not wholly clear how it should be modeled and what factors influence the behavior. The best predictive models we have of the behavior are information foraging and scent theories of web usage [5, 6, 7, 8, 25, 26]. Hence, in this paper, we report an empirical evaluation of information scent predictions based on an empirical study of web usage in a personalized search engine, called I-SPY [29, 30]. We find that these approaches do not make accurate predictions, prompting us to re-assess the cognitive basis of first-click behavior.

In the next section, we outline information scent approaches and the I-SPY system. Then, we sketch the empirical study of I-SPY. Next, we describe our empirical evaluation of information scent techniques and present the results found when these techniques are applied to the I-SPY data. In part response to the predictive failure found, we advance a general framework for assessing first-click behavior with a view to understanding what might be important in I-SPY. This same framework also helps us understand what it is that information scent approaches are trying to capture and how they may need to be modified to do better in the future.

2 Information Scent & I-SPY

Taxonomic studies have shown that users adopt several distinct types of behavior when using the Web, one of which is the specific goal-driven search for answers to specific questions [3, 23]. In the present paper, we are concerned with this type of directed search where a user has to answer a specific question by accessing a web page and does this by entering a query and selecting a link from a result list to meet that information need. The key contribution to be made by adaptive, personalized systems in this area is to increase the relevance ordering of result lists, so that the most relevant sites are in the initial positions of the result list. This research goal has become more acutely important with the emergence of the mobile Internet and the shrinking screen real estate available for presenting information.

Over the past few years, several tools and techniques have been developed for evaluating the usability of websites [2, 5, 6, 7, 8, 18, 26, 27]. In general, this work takes an information foraging approach, seeing human information seeking as being analogous to animals foraging for food [25]. This approach casts users as followers of information scents when web searching [2, 5, 6, 7, 8, 26]. The basic idea is that users will assess the distal content - namely, the page at the other end of the link - using proximal cues, the snippets of text or graphics that surround a link [5, 6, 7, 8]. By comparing these cues with their information goal, the user chooses the link that best meets their current goal, namely, the link with the highest information scent [2, 5, 6, 7, 8]. Two main flavors of information foraging have been advanced for usability assessment, the Cognitive Walkthrough for the Web [2] and the InfoScent Bloodhound Simulator [8]. To date, these approaches have been mainly used to provide (semi)-automated usability assessments of websites. However, they also make predictions about link-choices made by users in first-click behavior.

The Cognitive Walkthrough for the Web (CWW) is a theory-based inspection method used to evaluate how well a website supports users navigation and information search tasks [2]. CWW uses Latent Semantic Analysis (LSA) [19, 20] to calculate the information scent of a link given a specific user information need. In essence, it works by calculating the LSA-derived similarity between the users information goal (i.e., the query) and the text surrounding a given link. CWW has been shown to successfully predict uninformative/confusing links on analyzed websites [2].

The InfoScent Bloodhound Simulator [8] is an automated analysis system that examines the information cues on a website and produces a usability report. Bloodhound uses a predictive modeling algorithm called Web User Flow by Information Scent (WUFIS) that relies on information retrieval techniques (i.e., TF.IDF analyses) and spreading activation to simulate user actions based on their information needs [6, 8]. In this paper, for reasons of space, we concentrate on CWW rather than Bloodhound. It should be pointed out that CWW does better than Bloodhound.

We were interested in applying these approaches to a developed adaptive system, the personalized search engine I-SPY [29, 30]. I-SPY is an example of an adaptive information retrieval system. See, Micarelli & Sciarrone [22] and Pierrakos et al, [24] for other related work on adaptive information filtering and Web personalization.

I-SPY implements an adaptive collaborative search technique that enables it to selectively re-rank search results according to the learned preferences of a community of

users. Effectively I-SPY actively promotes results that have been previously favored by community members during related searches so that the most relevant results are top of the result list [29]. I-SPY monitors user selections or *hits* for a query and builds a model of query-page relevance based on the probability that a given page will be selected by the user when returned as a result to a specific query [1, 13, 29, 30].

I-SPY has previously been shown to be capable of generating superior result rankings, based on its collaborative model of page relevance [1, 13]. For instance, we know from the results of a live-user trial, designed to compare the search effectiveness of I-SPY users against a control group, that I-SPY's promoted results are likely to be valuable to searchers. This study provided us with access to comprehensive search logs reflecting detailed search behavior information including the queries submitted by control and test groups, the results returned and promoted, the results selected, and their positions when selected. If the information scent approaches accurately capture user behavior we should find that the results chosen by people are indeed those with the highest information scent.

3 Evaluating I-SPY

In this section we describe key aspects of the I-SPY evaluation. Although we do not evaluate the I-SPY system in this paper, we have included relevant details regarding the I-SPY evaluation to illustrate the environment and conditions in which we are attempting to evaluate information scent predictions for first-click behavior. For further details regarding the I-SPY evaluation see, [1, 14, 15, 29].

The I-SPY evaluation took place over two separate sessions and involved asking two separate groups of 45 and 47 Computer Science students, to answer a series of 25 questions on topics in Computer Science and Artificial Intelligence. In the first session, the I-SPY collaborative search function was disabled so the results presented to participants were drawn from a Meta search engine (using Google, AllTheWeb, Wisenut and HotBot). The students taking part in the first session served as a control group against which to judge the students taking part in the second session, where I-SPY's collaborative search function was enabled. When each person used I-SPY, they entered one or more query terms and were presented with a list of up to 20 results; consisting of a result number/rank, a title and a description/summary (or blurb).

A substantial amount of web search behavior data was generated from the I-SPY experiment. A total of 811 distinct queries were logged with 10,445 unique pages being returned in result lists. From these lists, a total of 427 unique pages were clicked on by users. All of this information was collected and archived for analysis in assessing the information foraging approaches to which we now turn.

4 Does CWW's Information Scent Predict First-Click Behavior

Cognitive Walkthrough for the Web [2] is a usability inspection technique for assessing information search tasks that uses Latent Semantic Analysis (LSA) [19, 20]. CWW calculates the information scent of a given piece of link text by finding its similarity (as computed by LSA) to the query terms used (or more commonly an elaborated description of the query). LSA captures patterns of co-occurrence between

words based on a text corpus analysis, in its most commonly used form, of general reading up to first-year college level in the US. LSA has been shown to predict some aspects of language comprehension and priming in various cognitive tasks [10, 11, 17, 19, 21]. In our analysis of the I-SPY data we looked at the information scent values computed from LSA for long/short queries compared against each and every piece of link text returned in the results lists (see Church, Keane & Smyth [9], for details).

4.1 Method

Data & Analysis Procedure. The analysis was performed on the data common to both I-SPY sessions. Limiting our dataset to just queries and urls common to both sessions of the I-SPY experiment provided us with valuable before and after ranking information as well as key web search behavior details. This common dataset consisted of 132 valid distinct queries and 2,571 results/urls.

Queries and results were sub-classified. For queries, we had short queries (the actual terms input by users) and long queries (an elaboration of the original question posed to the user). For results, we distinguished between (i) the title of the link alone, (ii) the blurb text alone (i.e., the summary text given back by the search engine) and (iii) both title and blurb. All six possible pairings for every unique query and result (2,571 distinct test items) were submitted to the LSA website [20]. In all, it took roughly 6 days of computing time to gather these scores. The short query - title case revealed the best results of all six possible pairings. One of the problems with LSA is that it does not contain many of the specialist terms used in Computer Science (we would assume that this is a general problem that would attach to any specialist domain). Hence, we filtered out those queries-result pairs that had large numbers of unknown terms to LSA (reduced set to 97 queries).

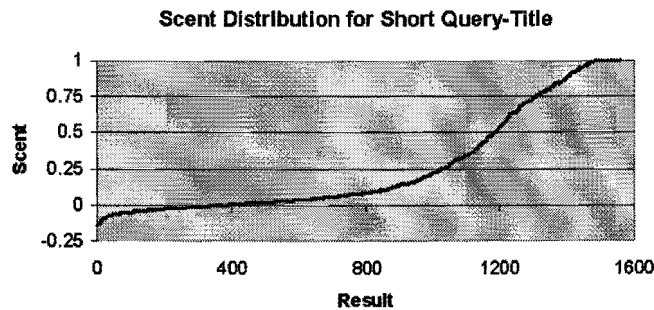


Fig. 1. Plot of the Scent Distribution for the Short Query-Title Pair in which the Results are Reordered by Scent

4.2 Results & Discussion

Overall, CWW does not do a good job of predicting the links clicked on by users in the I-SPY study. The correlations found are weak and, more damagingly, appear to mainly relay simple string-matching between the query and result.

Properties of the CWW Scent Values. Most of the scent values generated by LSA were low: the minimum score was -0.14 , the maximum was 1 ($M = 0.26$, $SD = 0.35$). Figure 1, above, shows the distribution of scent values for the short query-title pairing. In general, LSA generates high scent values when the text it receives has a high percentage of word-to-word matches and very few word-to-word mismatches (see our later analysis using string matching).

Does CWW’s Information Scent Predict Link Choice ? The crucial question for CWW is whether its scent values predict the pages chosen by people in the study. To carry out this evaluation we extracted a subset of I-SPY data that included only the common hit data (i.e., links that were chosen by users, all of whom entered an identical query with the same question in mind). This set consisted of 110 distinct queries and 1,218 chosen links. If CWW predicts people’s link choice then the hit score (i.e., the number of people choosing a given url) should correlate with its scent value. Unfortunately, this correlation is low. Table 1 shows the correlations between the short query and the three possible versions of the result (as title alone, blurb alone and title and blurb together). The correlations for long queries were worse, all ≤ 0.04 .

Table 1. Correlations between the Information Scent of the Short Query - Result Pair and the Hit Score

Result Type	Correlation	Classification
Short Query - Title	0.10	Very Weak
Short Query - Blurb	0.11	Very Weak
Short Query - Title + Blurb	0.13	Very Weak

Another perspective on this data can be gleaned from Figure 2 below, which shows scent values plotted by their hit scores for the best pairing (i.e., the short query – title + blurb). The most obvious conclusion from the graph is that many high-scented links have low hit scores and some low-scented links have high hit scores.

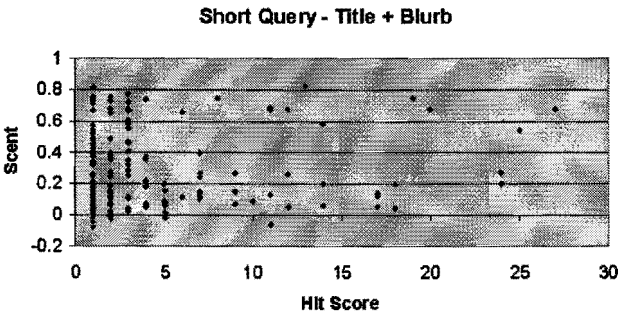


Fig. 2. Plot of the Scent Value by Hit Score for the Most Highly Correlated Query-Result Pair

One of the problems with this analysis is that some of the queries contain terms that are not in LSA's corpus¹. A fairer test would be to perform the same analysis for query-result pairs in which all the terms were known by LSA. However, even after removing all unsupported query-result pairs, the correlations did not improve appreciably, the highest being still low ($r = 0.15$).

Finally, to give CWW the best possible chance, we picked the most highly correlated cases (from the query-result matrix) to see what the best possible correlation could be. Figure 3 shows the result of this analysis. Note that each of the points here represent the scent value x hit score of a query-result pairing where that pairing could be any of the 6 possibilities (e.g., short query-title, long query-blurb, etc). In this best case, the correlation is better and moderate ($r = 0.5$). However, this good news must be tempered by the size and selective nature of the dataset.

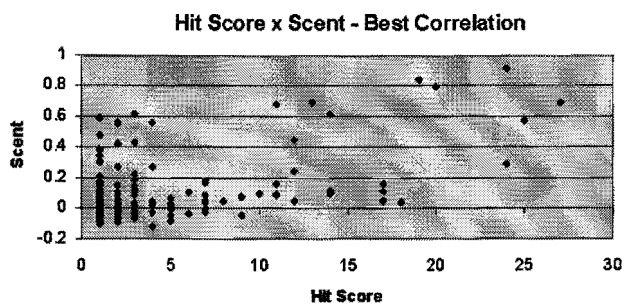


Fig. 3. Plot of the Scent Value by Hit Score for the Best-fit Correlations

Overall, we have to conclude that CWW can only generate reasonable predictions for a limited subset of queries (i.e., those for which it has a vocabulary) and even then it is not clear which one of 6 possible flavors of query-result pairing will best predict hits. This conclusion of limited appeal is further overshadowed by the possibility that CWW mainly appears to succeed by term matching.

Does CWW Succeed by Term Matching ? On the face of it, CWW's highest scent values seem to rely on string matching. To test this hypothesis, we applied a term matching model to the original dataset, using Tversky's Contrast Model of Similarity [32]. Tversky's model sees similarity as being based on the common and distinctive features of two items. The similarity between two objects a and b can thus be defined as,

$$S(a, b) = \theta f(A \cap B) - \beta f(A - B) - \alpha f(B - A) \quad (1)$$

where A represents the set of terms associated with a , B represents the set of terms associated with b , $A \cap B$ is the set of terms common to a and b , $A - B$ represents the set of features distinct in a , $B - A$ represents the set of features distinct in b . In this

¹ In these cases, LSA returns a value of N/A for the terms not in its corpus.

formula, f is usually a count of the features and θ , β and α are usually 1 or 0 (creating a family of models where one or another of the components can be cancelled out).

We then looked at the correlations between the scores produced by variants of the contrast model and CWW's scent values for the same items. We found that a simple term-matching model (i.e., the contrast model using only the common features component) was moderately-highly correlated with the CWW scent values ($r = 0.6$). Given that term matching is simpler and does not encounter the same vocabulary problems as CWW, it would appear to offer a better basis for predicting first-click behavior.

5 Bloodhound's Predictions for First-Click Behavior

We have carried out a similar analysis of the same data using the techniques developed in the InfoScent Bloodhound Simulator [8]. For reasons of space we cannot report these results here, suffice to say that the correlations are considerable worse than those found in CWW (see Church, Keane & Smyth [9], for details).

6 A Framework for Assessing First Clicks

Given our findings it is hard to escape the conclusion that, from a predictive perspective, we know very little about the basis of first-click behavior. As such, it would appear to be a good idea to step back from the problem and consider the main cognitive components of the context in which this task is being carried out. Card et al [5] have previously advanced a problem space of user's browsing behavior. However, we feel that their analysis is at a too fine-grained level to help us in this case and, maybe, really just provides us with a language with which to describe user behavior rather than a theory of the main parameters that impact that behavior. Therefore, in this section, we attempt to outline a general framework for understanding first-click behavior.

Broadly speaking, we can distinguish between the parts of the first click task that are represented in the user's head and those that are represented textually in the computer. The relevant data on the computer side are easily characterized and inspected. They include: the explicit question posed (if elaborated), the specific query terms used, the result lists returned, the ordering of those results, the distal pages to which these links refer and so on. On the human side, the relevant components are less easily characterized and not easily inspected. They include: the users' mental representation of the question, query and results, the user's background knowledge about the domain of the question, the user's knowledge of natural language, the user's knowledge of what ordered results entail, the user's similarity function for matching his/her information need to the presented result, the strategies the user normally employs when searching result lists, knowledge of previous searches and so on.

The key problem is that we do not have good techniques for acquiring and characterizing the knowledge that is brought to bear by users in choosing a link from a set of returned results. In theoretical terms, we need a well-developed cognitive model of this behavior. In practical terms, we need good proxies for this knowledge based on some analysis of the textual data we can explicitly enumerate in the task. In this sense a lot of the work to date can be characterized as proxies of varying goodness.

6.1 Some Proxies for User Knowledge

The usability methods employed here and many of the methods used to personalize and relevance-rank search engine outputs basically use some analysis technique that tries to approximate what people want using explicit data from the web context.

Link-Structure Analysis. Techniques that hinge on recommendation by analyzing link structure e.g., Google [4], essentially work on the assumption that the authority/relevance choices of a community of web-page builders, as indicated by their established links, will parallel the authority/relevance required by someone searching for a resource. The link structure created by the community is a proxy for the relevance ordering of the searching user.

Community-based Hits Analysis. Similarly, the techniques used by I-SPY, which hinge on analyzing the query-result choices of community users, also work on the assumption that what was good for others will be good for you. I-SPY's success is based on the closeness of this proxy to what the user is doing; using other people's choices to predict a new user's choice. In this respect, it is important to point out that I-SPY's builders assume that the community will be in some way representative of the user. This type of representative assumption is a familiar foundation for many approaches to lazy learning [31] and the degree to which it stands up in the context of I-SPY will depend largely on the focus of a particular community of searchers.

Corpus Analysis. CWW basically uses corpus analysis, based on LSA, to approximate a model of people's background knowledge for the words they use. On the face of it, this looks like a sound idea. But, other research has shown that LSA is not good at finding deep semantic similarity [12, 16]. This is exactly what our empirical analysis shows up. First, CWW fails because we fall off the edge of LSA's word knowledge (with specialist terms). Second, its generalization over word meanings is not powerful enough to be a good proxy to human knowledge.

Term-Frequency Analysis. Bloodhound makes heavy use of term frequency analysis in order to provide a proxy to people's knowledge of the domain. Our empirical studies show that varying the set of pages over which these values are computed (i.e., the domain) do not have a significant impact on the goodness of its predictions. It is hard to escape the conclusion that such term-frequency analyses are a poor proxy, on their own, for characterizing user knowledge.

7 Conclusions

Overall there are some positive and some negative conclusions to be made from the empirical analysis we have carried out here. Taking the bad news first, it is clear that current approaches using information scent do not do a good job of predicting the first clicks users make when presented with various results lists. In other words, we still need a good user model for this key behavior in web searching.

Happily, there are also a number of pieces of good news that we can take from this work. First, we outlined a methodology for the empirical evaluation of web search behavior. Second, we have shown that there are limitations to current information foraging theory that can be used productively to guide future theorizing. Third, with our presented framework, we have gained some perspective on the general nature of

first-click behavior. Fourth, we have seen that community-based hits analysis provides a reasonable proxy for first-click behavior, thus suggesting a fruitful direction for future work to characterize this behavior.

8 Acknowledgements

We thank the I-SPY research group at the University College Dublin, namely Jill Freyne, Evelyn Balfé, Peter Briggs and Maurice Coyle, who provided us with data from the I-SPY experiment. This work was funded by grants to the second and third authors from Science Foundation Ireland under Grant No.03/IN.3/I361.

9 References

1. Balfé, E., Smyth, B.: Case-Based Collaborative Web Search. In: Proceedings of the European Conference on Case-Based Reasoning, ECCBR'04, Springer (2004) Madrid, Spain.
2. Blackmon, M.H., Polson, P.G., Kitajima, M., Lewis, C.: Cognitive Walkthrough for the Web. In: Proceedings of the CHI 2002, ACM Press (2002) 463-470.
3. Broder, A.: A Taxonomy of Web Search. SIGIR Forum 36(2) (2002).
4. Brin, S., Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Proceedings of the 7th International World Wide Web Conference, (1998) 107-117 Brisbane, Australia.
5. Card, S.K., Pirolli, P., Van Der Wege, M., Morrison, J.B., Reeder, R.W., Schraedley, P.K., Boshart, J.: Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. In: Proceedings of CHI 2001, ACM Press (2001) 498-505 Seattle, WA.
6. Chi, E.H., Pirolli, P., Chen, K., Pitkow, J.: Using Information Scent to Model User Information Needs and Actions on the Web. In: Proceedings of CHI 2001, ACM Press (2001) 490-497 Seattle, WA.
7. Chi, E.H., Pirolli, P., Pitkow, J.: The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site. In: Proceedings of CHI 2000, ACM Press (2000) 161-168 The Hague, The Netherlands.
8. Chi, E.H., Rosien, A., Supattanasiri, G., Williams, A., Royer, C., Chow, C., Robles, E., Dalal, B., Chen, J., Cousins, S.: The Bloodhound Project: Automating Discovery of Web Usability Issues using the InfoScent™ Simulator. In: Proceedings of CHI 2003, ACM Press (2003) Fort Lauderdale, FL.
9. Church, K., Keane, M.T., Smyth, B.: Evaluating Cognitive & User Models of "First-Click Behavior" in a Personalized Search Engine. User Modeling and User-Adapted Interaction, *Submitted*.
10. Connell, L., Keane, M.T.: PAM: A Cognitive Model of Plausibility. In: Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society, Erlbaum (2003) Hillsdale, NJ.
11. Connell, L., Keane, M.T.: What Plausibly Affects Plausibility. Concept Coherence and Distributional Word Coherence as Factors Influencing Plausibility Judgments. *Memory & Cognition* (2004).
12. French, R., Labiouse, C.: Four Problems with Extracting Human Semantics from Large Text Corpora. In: Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society, Erlbaum (2002) Hillsdale, NJ.
13. Freyne, J., Smyth, B.: Collaborative Search: A Live User Trial. In: Proceedings of the 26th European Conference on IR Research, ECIR'04, (2004) Sunderland, UK.

14. Freyne, J., Smyth, B.: An Experiment in Social Search. In: Proceedings of the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, AH-04, (2004) Eindhoven, The Netherlands.
15. Freyne, J., Smyth, B., Coyle, M., Balfe, E., Briggs, P.: Further Experiments on Collaborative Ranking in Community-Based Web Search. *AI Review: An International Science and Engineering Journal*, *In Press*.
16. Glenberg, A.M., Robertson, D.A.: Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language*, 43 (2000) 379-401.
17. Kintsch, W.: Predication. *Cognitive Science*, 25 (2001) 173-202.
18. Kitajima, M., Blackmon, M.H., Polson, P.G.: A Comprehension-based Model of Web Navigation and Its Application to Web Usability Analysis. In *People and Computers XIV*, Springer (2000) 357-373.
19. Landauer, T.K., Dumais, S.T.: A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104 (1997) 211-240.
20. LSA Website: <http://lsa.colorado.edu>
21. Lund, K., Burgess, C., Atchley, R.A.: Semantic and Associative Priming in High-Dimensional Semantic Space. In: Proceedings of the 17th Annual Conference of the Cognitive Science Society, Erlbaum (1995) 660-665 Hillsdale, NJ.
22. Micarelli, A., Sciarrone, F.: Anatomy and Empirical Evaluation of an Adaptive Web-Based Information Filtering System. *User Modeling and User-Adapted Interaction*, 14(2-3) (2004) 159-200.
23. Morrison, J.B., Pirolli, P., Card, S.K.: A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions. In: Proceedings of CHI 2001, ACM Press (2001) 163-164 Seattle, WA.
24. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.D.: Web Usage Mining as a Tool for Personalization: A Survey. *User Modeling and User-Adapted Interaction*, 13(4) (2003) 311-372.
25. Pirolli, P., Card, S.K.: Information Foraging. *Psychological Review*, 106(4) (1999) 643-675.
26. Pirolli, P., Fu, W.-T.: SNIF-ACT: A Model of Information Foraging on the World Wide Web. In: Proceedings of the 9th International Conference on User Modeling, (2003) Johnstown, PA.
27. Polson, P.G., Lewis, C., Rieman, J., Wharton, C.: Cognitive Walkthroughs: A Method for Theory-Based Evaluation of User Interfaces. *International Journal of Man-Machine Studies*, 36 (1992) 741-773.
28. Salton, G.: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley (1989).
29. Smyth, B., Balfe, E., Briggs, P., Coyle, M. and Freyne, J.: Collaborative Web Search. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence, IJCAI-03, Morgan Kaufmann (2003) 1417-1419 Acapulco, Mexico.
30. Smyth, B., Freyne, J., Coyle, M., Briggs, P., Balfe, E.: I-SPY: Anonymous, Community-Based Personalization by Collaborative Web Search. In: Proceedings of the 23rd SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer (2003) 367-380 Cambridge, UK.
31. Smyth, B., Keane, M.T.: Adaptation-Guided Retrieval: Questioning the Similarity Assumption in Reasoning. *Artificial Intelligence*, (102)2 (1998) 249-293.
32. Tversky, A.: Features of Similarity. *Psychological Review*, 84(4) (1997) 327-352.

Evaluating Intelligent Tutoring Systems with Learning Curves

Brent Martin and Antonija Mitrovic

Intelligent Computer Tutoring Group
Department of Computer Science, University of Canterbury
Private Bag 4800, Christchurch, New Zealand
{brent,tanja}@cosc.canterbury.ac.nz

Abstract. The evaluation of Intelligent Tutoring Systems, like any adaptive system, can have its difficulties. In this paper we discuss the evaluation of an extension to an existing system that uses Constraint-Based Modelling (CBM). CBM is a student modelling technique that is rapidly maturing, and is suited to complex, open-ended domains. A problem with complex domain models is their large size, necessitating a comprehensive problem set in order to provide sufficient exercises for extended learning sessions. We have addressed this issue by developing an algorithm that automatically generates new problems directly from the domain knowledge base. However, evaluation of this approach was complicated by the need for a lengthy (and therefore uncontrolled) study as well as other unavoidable differences between the control and experimental systems. This paper presents the evaluation and discusses those issues, and the way in which we used learning curves as a tool for comparing disparate learning systems.

1 Introduction

Constraint-Based Modelling (CBM) [7] is an effective approach that simplifies the building of domain and student models in Intelligent Tutoring Systems (ITS). We have used CBM to develop SQL-Tutor [4], an ITS for teaching the SQL database language. SQL-Tutor supports learning in three ways: by presenting feedback when students submit their answers, by controlling problem difficulty, and by providing scaffolding information. Students have shown significant gains in learning after as little as two hours of exposure to this system [5].

SQL-Tutor contains a list of problems, from which one is selected that best fits the student's current knowledge state. In extended sessions with the tutor the system may run out of suitable problems. We have attempted to overcome this by developing a problem generator that uses the domain model to build new problems that fit the student model, and a problem selection algorithm that is tied more closely to the student model.

We evaluated the new algorithm against the existing SQL-Tutor system in a classroom setting, but had difficulty because there were other differences in the system besides what we were testing. Also, the evaluation was over an extended period in an uncontrolled environment, which made comparisons in student performance difficult. We were therefore unlikely to see any difference in outcome using purely subject pre- and post-testing.

In this paper we briefly introduce SQL-Tutor and the extensions to it. We then describe the experiment and its evaluation. We discuss the use of learning curves to analyse the performance of the two systems, and introduce a new measure, that of the initial learning rate. We compare this new measure to the usual ones of slope and intercept, and discuss why we think it is valid. Finally, we conclude why we think the initial learning rate may be a better measure for determining how much learning went on while students used each system.

2 SQL-Tutor

SQL-Tutor [4] teaches the SQL database query language to second and third year students at the University of Canterbury, using Constraint-Based Modelling (CBM). This approach models the domain as a set of state constraints, of the form:

If <relevance condition> is true for the student's solution,
THEN <satisfaction condition> must also be true

The relevance condition of each constraint is used to test whether the student's solution is in a pedagogically significant state. If so, the satisfaction condition is checked, which tests for valid constructs (syntactic constraints) or compares the solution to an ideal solution (semantic constraints). If a constraint succeeds, no action is taken; if it fails, the student has made a mistake and appropriate feedback is given. CBM has advantages over other approaches such as model tracing [1] in that the model need not be complete and correct in order to function. Further, it is well suited to domains where the number of alternative solutions is large or the domain is open-ended.

Ohlsson [7] does not impose any restrictions upon how constraints are encoded, and/or implemented. In SQL-Tutor we initially represented each constraint by a LISP fragment, supported by domain-specific LISP functions. In later versions we have used a pattern-matching algorithm designed for this purpose [3], for example:

```
(147
"You have used some names in the WHERE clause that are not from
this database."
(match SS WHERE (?* (^name ?n) ?*))
(or (test SS (^valid-table (?n ?t))
(test SS (^attribute-p (?n ?a ?t))))
"WHERE")
```

The above constraint checks for valid table and attribute names in the WHERE clause of the student's SQL statement. It is relevant if any names exist in the WHERE clause, and satisfied if each of these is either a valid table name or a valid attribute name.

The constraint language makes all of the logic for determining whether or not the constraint is satisfied transparent to the system, since it consists only of pattern matching and logical combination. Functions, such as “^valid-table” in the above example, are simply macros defined in the same language. We have used this property to develop a problem solver that can generate correct solutions from student

attempts by extracting (and unifying) the valid match patterns from each satisfied constraint and the *invalid* patterns from violated constraints. The algorithm then corrects the invalid fragments by unifying them against matched patterns of the *ideal* solution, and then combines the resulting solution fragments.

Unlike repair theory [8], we make no claim that this algorithm is modelling human behaviour. However, it has the advantage that a failed constraint means that the construct involved is *definitely wrong*: we do not need to try to infer where the error lies, so our algorithm does not suffer from computational explosion. For further details on this algorithm and the constraint language refer to [3].

3 Generating New Problems

In the original version of SQL-Tutor the next problem is chosen based on difficulty, plus the concept the student is currently having the most trouble with. The constraint set is first searched for the constraint that is being violated most often. Then the system identifies problems for which this constraint is relevant. Finally, it chooses the problem from this set that best matches the student's current proficiency level. However, there is no guarantee that an untried problem exists that matches the student model: there may be no problems for the target constraint, or the only problems available may be of unsuitable difficulty. Further, since the constraint set is large (over 650 constraints), many problems are needed merely to cover the domain. Ideally there should be many problems per constraint, in various combinations. In SQL-Tutor there is an average of three problems per constraint and only around half of the constraint set is covered. A consequence of this is that the number of new constraints being presented to the student tapers off as the system runs out of problems.

The obvious way to address this limitation is to add more problems. However, this is not an easy task. There are over 650 constraints, and it is difficult to invent problems that are guaranteed to cover all constraints in sufficient combinations that there are enough problems at a large spread of difficulty levels. To overcome this we have developed an algorithm that generates new problems from the constraint set. It uses the constraint-based problem solver described in Section 2 to create novel SQL statements, using an individual constraint (or, possibly, a set of compatible constraints) to provide a set of SQL fragments as a starting point. These are then "grown" into a full SQL statement by repeatedly splicing them together and unifying them against the syntactic constraint set until no constraints are violated. This new SQL statement forms the ideal solution for a new problem. The human author need only convert the ideal solution into a natural language problem statement, ready for presentation to the student.

We used the problem generation algorithm create a single problem per constraint, giving around 800 potential ideal solutions (note that the experimental version had an extra 250 constraints added). We then chose the best of these and converted them into natural language problem statements. On completion we had a new problem set of 200 problems, which took less than a day of human effort to build. Furthermore, when we plotted the number of new constraints applied per problem presented to a student, the point at which new problems failed to introduce any new concepts (i.e. previously unseen constraints) rose from 40 problems to 60, indicating that the new problem set increased the length of time that a student could fruitfully engage with the system.

The experimental system also used a different problem selection mechanism. In the control system, a new problem is selected by finding the constraint the student is currently violating most often and selecting the problem whose (static) difficulty rating is closest to the student's current proficiency level. In the experimental system we calculated the difficulty of each problem dynamically by computing its static difficulty from the number of constraints (and their complexity) relevant to it, and then added further difficulty for relevant constraints that the student is currently violating, and more still for constraints that the student has not yet encountered. Thus each problem is compared to the student's current knowledge. Once difficulties have been calculated for all problems, the one that is closest to the student's current proficiency is selected.

4 Evaluation

The motivation for Problem Generation was to reduce the effort involved in building tutoring systems by automating one of the more time-consuming functions: writing the problem set. Three criteria must be met to achieve this goal: the algorithm must work (i.e. it must generate new problems); it must require (substantially) less human involvement than traditional problem authoring; and the problems produced must be shown to facilitate learning to at least the same degree as human-authored problems. The first two were confirmed during the building of the evaluation system: the algorithm successfully generated problems, and the time taken to author the problem set was much less than would have been required for human authoring alone.

Additionally, if the method works, it should be possible to generate large problem sets, which will have the benefit of greater choice when trying to fit a problem to the user's current student model. We might therefore expect that given a suitable problem selection strategy, a system using the generated problem set would lead to faster learning than the current human-authored set, because we are better able to fit the problem to the student.

SQL-TUTOR was modified for this purpose and evaluated for a six-week period. The subjects were second year University students studying databases. The students were broken into three groups. The first used the current version of SQL-TUTOR, i.e. with human-authored problems. The second group used a version with problems generated using the algorithm described. The third group used a version containing other research (student model visualisation) that was not relevant to this study. Before using the system, each student sat a pre-test to determine their existing knowledge and skill in writing SQL queries. They were then free to use the system as little or as often as they liked over a six week period. Each student was randomly assigned a "mode" that determined which version of the system they would use. At the conclusion of the evaluation they sat a post-test.

When the study commenced, 88 students had signed up and sat the pre-test, giving sample sizes of around 30 per group. During the evaluation this further increased as new students requested access to the system. At the conclusion of the study some students who signed up had not used the system to any significant degree. The final groups used for analysis numbered 24 (control) and 26 (experimental) students each. The length of time each student used the system varied greatly from not using it at all

to working for several hours, with an average of 2½ hours. Consequently, the number of problems solved also varied widely from zero to 98, with an average of 25.

There are several ways we can measure the performance of the system. First, we can measure the means of the pre-test and post-test to determine whether or not the systems had differing effects on test performance. Note, however, that with such an open evaluation as this it is dangerous to assume that differences are due to the system, since use of the system may represent only a portion of the effort the student spent learning SQL. Nevertheless, it is important to analyse the *pre-test* scores to determine whether the study groups are comparable samples of the population. This was found to be the case. There was similarly no significant difference in post-test scores, as we might expect.

Second, we can plot the reduction in error rates as the student practices on each constraint, or the “learning curve”. Each student’s performance when measured this way should lead to an exponential curve or so-called “Power law” [2, 6], which is typical when each underlying object being measured (in this case a constraint) represents a concept being learned. The steepness of this curve is an indication of the speed with which the student, on average, is learning new constraints. Since each constraint represents a specific concept in the domain, this is an indication of how quickly the student is learning the domain. We can then compare this learning rate between the two groups.

Finally, we can look at how difficult the students found the problems. This is necessary to ensure that the newly generated problems did not negatively impact problem difficulty (either by being too easy or too hard). There are several ways we can do this. First, we can measure how many attempts the student took on average to solve a problem and compare the means for the control and test groups. Second, students were permitted to abort the current problem and were asked to cite one of three reasons: it was too easy, it was too hard or they wanted to try a problem of a different type. If the proportion of problems aborted rises, or the ratio of “too hard” to “too easy” problems is further from 1:1 than the control group, we might conclude that problem difficulty has been adversely affected.

In this study, we measured all of the above. We used the software package SPSS to compare means and estimate power and effect size, and Microsoft Excel to fit power curves. We measured problem difficulty both subjectively and objectively: we obtained subjective results by logging when students aborted a problem and recording their reason. Any significant difference in the ratio of “too easy” to “too hard” responses would suggest we have adversely affected problem difficulty. Further, the percentage of problems aborted should not rise significantly. Next, we measured the number of attempts taken to solve each problem, and the time taken on each attempt. There was no significant difference in any of these comparisons. In other words, the students appeared to have found the level of difficulty of each problem about the same for both systems.

4.1 Learning Curves

Since the general tests failed to show any difference between the two groups, we turned to learning curves as a means of evaluating more closely what was occurring *while the system was being used*. We observed the learning rate for each group by plotting the proportion of constraints that are violated for the n th problem for which

this constraint is relevant. This value is obtained by determining for each constraint whether it is correctly or incorrectly applied to the *n*th problem for which it is relevant. A constraint is correctly applied if it is relevant at any time during solving of this problem, and is *always* satisfied for the duration of this problem. Constraints that are relevant but are violated one or more times during solving of this problem are labelled erroneous. The value plotted is the proportion of all constraints relevant to the *n*th problem that are erroneous.

If the unit being measured (constraints in this case) is a valid abstraction of what is being learned, we expect to see a “power curve”. We fitted a power curve to each plot, giving an equation for the curve. Note that as the curve progresses learning behaviour becomes swamped by random erroneous behaviour such as slips, so the plot stops trending along the power curve and levels out at the level of random mistakes. This is exacerbated by the fact that the number of constraints being considered reduces as *n* increases, because many constraints are only relevant to a small number of problems. We therefore use only the initial part of the curve to calculate the learning rate. Fig. 1 shows such plots, where each line is the learning curve for the entire group on average, i.e. the proportion of constraints that are relevant to the first problem that are incorrectly applied *by any student in the group*. The cut-off was chosen at *n*=5, which is the point at which the power curve fit for both groups is maximal. Note that learning curves are also exhibited when the data is plotted for a single student, although there is large variance between faster and slower learners, and the quality of the curves is often lower due to the small amount of data available. Similarly, plotting learning curves on a per-constraint basis (averaged over all students) gives power laws where the slope indicates the ease with which this constraint is learned, which is an indication of the quality of the constraint. For example, constraints that span more than one concept will produce a poor curve.

Both plots exhibit a very good fit to a power curve. The equations give us the Y intercept and slope for a log-log graph of constraint performance. In this case the experimental group had a Y intercept that was around *twice* that for the control group,

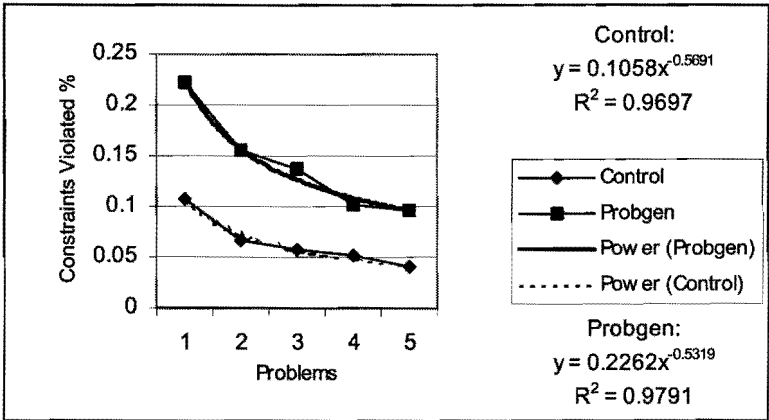


Fig. 1. Learning Performance

but a lightly lower slope (0.53 versus 0.57). Fig. 2 shows log-log plots for the same data.

4.2 Learning Curve Slope

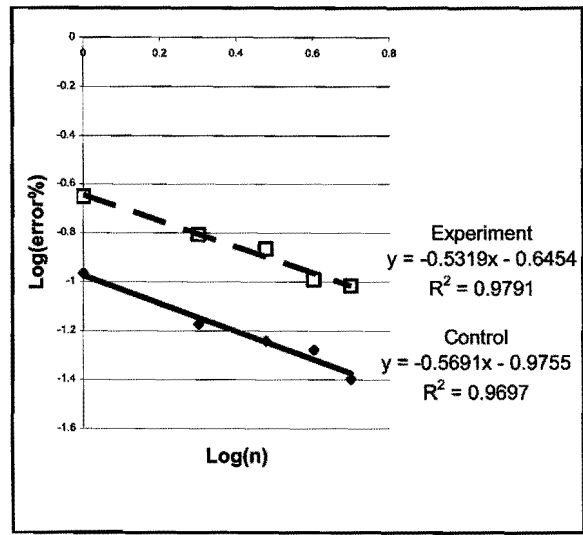


Fig. 2. Log-Log plots for the two groups.

The slope of a learning curve represents the *learning rate*, i.e. the rate at which the student decreased their errors with practice. There are several reasons why this might differ between the two groups: the students may have different average learning abilities; the constraints may represent domain concepts to a greater or lesser degree; the concepts may be introduced at more or less opportune times (and thus the students may be more or less receptive to learning the constraints at

that time).

The first difference (learner ability) was eliminated because the average pre-test scores were not significantly different for the two groups. Differences in the constraint sets were a possibility because the experimental system was a rewritten version of the control system using the new constraint representation: although the constraint set for the experimental group was based on that for the control group, a large number of modifications had been made, including the rewriting of many constraints that were always relevant, such that they were now only relevant in appropriate situations. New constraints were also added.

We tested the effect of these changes by recalculating the curves in two ways. First, we removed the data for all constraints that were trivially true (which occurred only in the control system's constraint set) and replotted the learning curve for the control group. Second, we took the raw student solutions from logs for the control group and evaluated them using the new constraint set. This latter method is a trade-off: it gives us the student's performance based on evaluation by the new constraints, but where the student still received feedback from the old constraint set.

In both cases the slope varied according to which method we used. Table 1 lists the results. The learning curve slope thus appears to be very sensitive to how the student's performance is being measured, and hence is not a suitable measure for comparing disparate systems.

Table 1. Learning curve slope for different constraint sets

Measuring method (Control Group)	Control Slope	Experiment Slope
Original constraint set	0.57	0.53
Exclude trivially true	0.31	0.53
Experimental constraint set	0.44	0.53

4.3 Y Intercept and Initial Learning Rate

The other measure of a learning curve is the Y intercept, which gives us the initial error rate. This alone is not a useful measure because it only indicates the student's initial performance, but does not show any effects of learning. However, the *slope* at $X=1$ (equal to the Y intercept multiplied by the log-log slope) does take learning into account: it indicates the raw improvement in performance achieved by a student between when they are first exposed to a constraint and when they have received feedback on it once. We therefore hypothesised that this is a better measurement of performance between disparate systems because it takes into account both the student's learning rate and the amount of unlearned material they are being exposed to as a percentage of the original constraint set. We expected that this calculation would be relatively invariant to changes in the constraint set, because it normalises out differences in the way the student's learning performance is measured.

Table 2 lists the results for the three ways that the students' performance was measured, i.e. using the original constraint, the same constraint set with trivial constraints excluded, and the new (experimental) constraint set. They show that, although the initial learning rate does differ depending on the constraint set used, the difference is much smaller than when the log-log slope is considered.

These results show that the slope at $X=1$, or initial learning rate, is almost twice as high for the experimental group. We checked for statistical significance by plotting learning curves for each individual student and comparing the means of the initial learning rate for the two groups. The difference was statistically significant at $\alpha=0.05$ ($p=0.01$). This suggests that the raw amount learned by the experimental group was higher than for the control group, which we attributed to either, or both, the increased number and range of problems available and an improved problem selection mechanism. Both of these could lead to a larger number of unknown constraints being presented to the student that were within their zone of proximal development [9]. Thus whereas the new problem set and selection method might not increase the student's learning rate, nevertheless it engages them in a larger volume of learning and therefore reduces the time taken to master the material.

Table 2. Initial slope for different constraint sets

Measuring method (Control Group)	Control Init. Slope	Exp. Init. Slope
Original constraint set	0.06	0.12
Exclude trivially true	0.05	0.12
Experimental constraint set	0.08	0.12

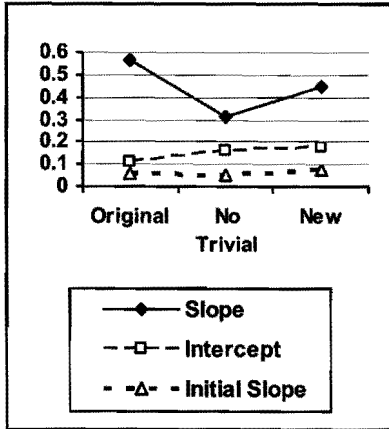


Fig. 3. Learning curve statistics.

only 10%, and is thus learning a greater volume of material at once. In both cases the ease with which the student learns the material is the same, but the amount they are learning varies widely. On the other hand, the Y intercept gives us the amount of learning being undertaken (i.e. what percentage of the presented material the student is trying to learn) but not the rate. However, the *initial* learning rate (slope at X=1) combines both the size of the learning task and the student's learning performance.

This new measure appears to be relatively invariant to the way student performance is measured, the issue in this case being differences in the constraint set. Fig. 3 plots how the measures of slope, Y intercept and initial slope vary with the constraint set used. Of the three, the initial slope varies the least. This is intuitively expected, and is illustrated in Fig. 4. These three curves are raw data for the control group, the same data multiplied by two, and again with a constant (0.5) added. The first two curves

5 Discussion

The two main issues with this study were how to measure learning differences between two dissimilar systems, and which measure to use to demonstrate differences that we were actually interested in. Learning rate (based on learning curve slope) was considered but discounted because it measures how fast the student is learning, but is relative to how their performance is being measured. Thus, it can be expected to vary between different systems. Further, it does not give any indication of the size of the learning task: the same numbers may be obtained for a student who knows 95% of the material they are initially presented with as one who knows

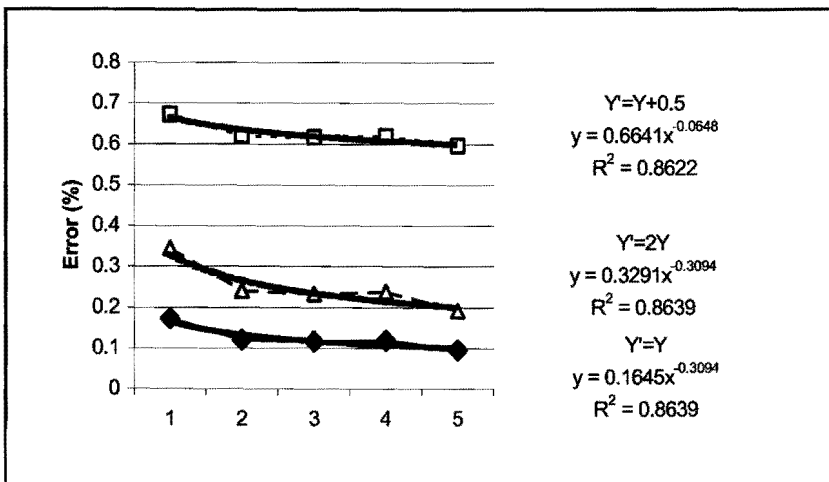


Fig. 4. Variations in power curve slope from artificial manipulation

have the same slope, even though the proportion of constraints being learned is twice as high for the second curve. However, the *initial* slope for the two curves has a ratio of 2:1, representing this effect. In contrast, the third curve is the same data with a constant amount added, which might represent a constraint set that includes trivially satisfied constraints. In this case the slope of the modified curve is dramatically lower (0.06 versus 0.3), even though the student's behaviour is unchanged. In contrast the initial learning rate is only slightly lower (0.43 versus 0.5).

6 Conclusions

This paper identified the problem of measuring student performance with intelligent tutoring systems when the systems being compared do not measure student performance in the same way. We showed that learning curves might still be used to compare such systems, but that the traditional measure of slope is not suitable because it varies with the method used to measure performance. We suggested a new statistic, that of the initial learning rate, which is produced by calculating the slope at $X=1$. We argued that this measure encompasses both the learning rate and the size of the learning task, and hence tells us more about the difference in performance between the two systems, because modifications to some aspects of a learning system (such as the problem set and the problem selection method) may alter the amount of material a student is learning at any time. We also showed that initial learning slope appears to be relatively robust against differences in the method used to measure the student's performance because it is normalised with respect to the student's initial performance.

Acknowledgement

This research was supported by the University of Canterbury grant U6430.

References

1. Anderson, J.R., Corbett, A.T., Koedinger, K.R., and Pelletier, R., *Cognitive Tutors: Lessons Learned*. Journal of the Learning Sciences, 1995. 4(2): pp. 167-207.
2. Anderson, J.R. and Lebiere, C., *The atomic components of thought*. 1998, MahWah, NJ: Lawrence Erlbaum Associates.
3. Martin, B. and Mitrovic, A. *Tailoring Feedback by Correcting Student Answers*. in *Fifth Int. Conf. on Intelligent Tutoring Systems*. 2000. Montreal: Springer. pp. 383-392.
4. Mitrovic, A. *Experiences in Implementing Constraint-Based Modeling in SQL-Tutor*. in *4th Int. Conf. on Intelligent Tutoring Systems*. 1998. San Antonio, Texas: Springer. pp. 414-423.
5. Mitrovic, A. and Ohlsson, S., *Evaluation of a Constraint-Based Tutor for a Database Language*. Int. J. Artificial Intelligence in Education, 1999. 10: pp. 238-256.
6. Newell, A. and Rosenbloom, P.S., *Mechanisms of skill acquisition and the law of practice*, in *Cognitive skills and their acquisition*, J.R. Anderson (ed.) 1981, Lawrence Erlbaum Associates: Hillsdale, NJ. pp. 1-56.
7. Ohlsson, S. *Constraint-Based Student Modeling*. in *NATO Advanced Research Workshop on Student Modelling*. 1991. Ste. Adele, Quebec, Canada: Springer-Verlag. pp. 167-189.
8. VanLehn, K., *On the Representation of Procedures in Repair Theory*, in *The Development of Mathematical Thinking*, H.P. Ginsburg (ed). 1983, Academic Press: New York. pp. 201-252.
9. Vygotsky, L.S., *Mind in society: The development of higher psychological processes*. 1978, Cambridge, MA: Harvard University Press.

Evaluation of Effects on Retrieval Performance for an Adaptive User Model

Hien Nguyen, Eugene Santos Jr., Qunhua Zhao, and Chester Lee

Computer Science and Engineering Department
University of Connecticut
191 Auditorium Road, U-155, Storrs, CT 06269-3155
{hien,eugene,qzhao,cllee}@cse.uconn.edu

Abstract. One of the challenging problems for evaluating the effectiveness of a user model with regards to retrieval performance is the absence of an evaluation method that offers the ability to compare with other existing approaches while assessing the new features offered by a user model. In this paper, we report our method of using collections, procedures and metrics from the information retrieval community to evaluate a cognitive user model which captures user intent to improve retrieval performance and adapts to a user's interests, preferences and context. Specifically, by starting with an empty user model for each query, we simulate the process of assessing the short-term effects of relevance feedback techniques in traditional information retrieval. By using a seed user model learned from relevance feedback, we assess both short and long-term effects on the entire search session. In this paper, we show how we can compare user modeling approaches by using the above method against a classic information retrieval approach, the Ide dec-hi, using CACM and Medline collections. This evaluation also helps analyze and address the strengths and weaknesses of our model and develops appropriate solutions.

1 Introduction

One of the challenging problems with evaluating an adaptive user model for information retrieval (IR) is the absence of an evaluation method that offers the ability to compare with other existing approaches in IR community while accessing the new, special features brought by the model. In the user modeling (UM) community, many researchers have explored the use of user models for improving retrieval performance [1, 2] and have evaluated the effectiveness of their user models on retrieval performance by using their own collections, tasks and procedures. Therefore, it is very difficult to compare them against different techniques, especially against the techniques used in the IR community. On the other hand, standard metrics, collections and procedures have been established and used in the IR community for decades to evaluate different retrieval techniques, especially the techniques that use relevance feedback (RF) and query expansion (QE) to improve retrieval performance [13]. However, the user model

created by using IR techniques such as RF and QE is short-lived. The model only affects a single query instead of the entire search session. In addition, these techniques assessed the retrieval performance in isolation. Therefore, using these procedures alone may not fully evaluate the special features that are created by long-lived user models.

In this paper, we report our evaluation with regards to retrieval performance for a cognitive user model which captures user intent for IR [14–16]. This is one important phase of an ongoing three-phase evaluation proposed in [10] in which we evaluate the correctness of the process of capturing user intent, the effectiveness of the user model on retrieval performance and user performance. The power of our method lies with its objectivity, inexpensiveness and comparability. The objectivity is reflected in using the collections and metrics, which do not depend on a particular set of users nor a set of parameters used in the adaptive system being tested. The procedures are lightweight and can be used for any other adaptive system in information retrieval. The comparability is achieved by using a standard procedure as a part of our evaluation, in which we simulate the traditional procedures used in the IR community. We seek to address two important questions: (1) How can we use collections, metrics and procedures from the IR community to evaluate our user model, especially its short-term and long-term effects? and (2) How does this evaluation help us analyze the overall effectiveness of our user model on user and system performance? Our user model captures user intent dynamically by analyzing information in retrieved relevant documents. Therefore, we compare our approach with the best traditional approach for RF, the *Ide dec-hi* approach using term frequency inverted document frequency weighting scheme (TFIDF) [13] on CACM and Medline collections.

This paper is organized as follows: We begin with a review of some important related work in IR and UM communities regarding the evaluation of a user model for IR. Next, we briefly present our approach. Then our evaluation method is presented, followed by the analysis of the results and our discussion. Finally, we present our conclusions and future work.

2 Related work

The main problems for evaluating the effectiveness of a user model for IR in terms of retrieval performance lie with the difficulty in comparing different approaches and the limitation of using this result to improve user performance. These problems are the results of little overlap between IR and UM communities when building user models for IR, as identified in [17].

In the IR community, user models have been created by using IR techniques such as RF and QE [18]. IR researchers have developed metrics, procedures and collections to assess the effectiveness of these two approaches for decades. Specifically, in the study done by Salton and Buckley [13], a common evaluation framework has been laid out to evaluate twelve different RF techniques, including *Ide dec-hi*. This framework offers the ability to compare different techniques with each other by using average precision at three specific recall points (0.25, 0.5 and

0.75) (we call this *three point fixed recall*). It also ensures that we assess a RF technique based on *new* information retrieved by using residual collections. A residual collection is created by removing all documents previously seen by a user from the original collection regardless of whether they are relevant or not; then the evaluation process is done using the reduced collection only. Some other techniques, such as freezing and test/control groups, are used to evaluate RF and QE techniques [20].

In the UM community, the common practice is to perform experiments with a particular set of users with and without the presence of a user model [5]. While this process is definitely needed to evaluate any adaptive system, it is expensive and the result depends on the particular group of users who participated in the experiments. Therefore, in order to better prepare for the experiments with real users, it is a good idea to first test a system within a simulated environment. By combining the results in the simulated and real environments, we can further analyze the outcomes from different perspectives. So far, most of the studies [1–3, 9] use two common metrics in IR: precision and recall [12]. These studies, unfortunately, use their own test collections and tasks; thus making any comparison difficult for current and future approaches.

3 Our user modeling approach

In our model, we capture, represent and use the information on *what* a user is currently interested in (the Interests); *how* a query needs to be constructed (the Preferences) and *why* the user dwells on a search topic (the Context) to modify a user's queries pro-actively.

Our user model architecture

We capture the Context, the Interests, and the Preferences aspects of a user's intent with a *context network* (C), an *interest set* (I), and a *preference network* (P). While previous efforts have focused on capturing just a single aspect, none of them have combined these three aspects in capturing user intent. A context network (C) is a directed acyclic graph (DAG) that contains *concept nodes* and *relation nodes*. Concept nodes are noun phrases representing the concepts found in retrieved relevant documents (e.g. "*computer science*"). Relation nodes represent the relations among these concepts. There are two relations captured: set-subset ("*isa*") and relate-to relations ("*related to*"). We construct C dynamically by finding a set of subgraphs in the intersection of all retrieved relevant documents. Each document is represented as a *document graph* (DG), which is also a DAG. We developed a program to automatically extract DG from text. We extracted noun phrases (NPs) from text using Link Parser [19]; these NPs will become concept nodes in a DG. The relations nodes are created by using three heuristic rules: *noun phrase heuristic*, *noun phrase-preposition phrase heuristic*, and *sentence heuristic*.

Each node in C is associated with its *weight*, *value* and *bias*, which are used by a spreading activation algorithm to reason about the new set I . In this algorithm, a concept that is located far from an observed interest concept will be of less

interest to the user. After we find the set of common subgraphs, we check to see if a subgraph is not currently in C , and add it accordingly. We ensure that the update will not result in a loop in C . As we can see from the representation of C , which contains the relations between concept nodes which represent potential goals that a user wants to explore. Therefore, it can be used to explain why a user is particularly interested in this concept based on its relations with more general/more specific/or related concepts.

Each element of I consists of an *interest concept* (a) and an *interest level* ($L(a)$). An interest concept refers to a concept a user is focusing on, and an interest level is any real value from 0 to 1 representing how much emphasis the user places on a concept. Based on the values of each interest concept found in C , we produce a rank ordering of the concepts to build I . Since a user's interests change over time, we incorporate a fading function to make the likely irrelevant interests fade away. We compute $L(a)$ after every query by: $L(a) = 0.5 * (L(a) + \frac{n}{m})$ with n as the number of retrieved relevant documents and m as the number of retrieved documents containing this concept a . If $L(a)$ falls below a threshold, a is removed from I .

We use a Bayesian network [7] to represent P because of its expressiveness, and ability to modeling uncertainty. There are 3 kinds of nodes in P : pre-condition (Pc), goal(G), and action (A) nodes. A user's query and the concepts contained in I are examples of Pc . An example of G is a tool called a filter that narrows down the search topics semantically or an expander that expands the search topics semantically. An example of A is a modified query. For each pre-condition node representing a user's current interest, its prior probability will be set as the interest level of the corresponding interest concept. The conditional probability table of each goal node is similar to the truth table of logical AND. Each G is associated with only one A . The probability of A is set to 1 if the tool is chosen and to 0, otherwise.

P is updated when a user gives feedback. The preference network adapts based on the observation of interactions between a user and our system. Two new preference networks are created; one of them contains a new tool labelled *filter*, and another contains a new tool labelled *expander*. The correction function calculates the probability of a new network that improves the user's effectiveness for both of the two new preference networks. The preference network is updated according to the one with higher probability. The calculation is used to determine the frequency that a tool helps in the previous retrieval processes. If the total number of retrieved relevant documents exceeds a threshold, the tool is considered helpful.

Integrating our user model into an IR system

- Given a user model $M=\{I, P, C\}$ and a query graph (QG) q . A QG is similar to a DG but is built from a user's query.
- Re-compute the values of interest concepts found in C by a spreading activation algorithm on C to construct I' .
- We set as evidence all interest concepts of I' found in P . Find a pre-condition node Pc representing a query in P which has associated query graph(QG)

that completely or partially matches against the given q . If such a node Pc is found, set it as evidence.

- Perform belief updating on P . Choose top n goal nodes from P with highest probability values. We call this set of goals as SG .
- For every goal node g in SG : If the query has been previously submitted and the user has used g , replace the original query subgraph with the graph associated with the action node of this goal. If the query has not been submitted before and g represents a filter: For every concept node q_i in the user's query graph q , we search for its corresponding node cq_i in C . For every concept a_i in I , we search for its corresponding node ca_i in C such that ca_i is an ancestor of cq_i . If such ca_i and cq_i are found, we add the paths from C between these two nodes to the modified query graph. It works similarly with an expander except that ca_i should be a progeny of cq_i .

The modified QG is sent to the search module where it is matched against each DG representing a record in our database. Those records that have the number of matches greater than a user-defined threshold are chosen and displayed to a user. A match between a QG q and a DG d_i is defined as $sim(q, d_i) = \frac{n}{2 \cdot N} + \frac{m}{2 \cdot M}$ in which n, m are the number of concepts and relation nodes of q found in d_i , respectively. N, M are the total number of concept and relation nodes of q . Two relation nodes are matched if and only if at least one of their parents and one of their children are matched.

4 Evaluation method

4.1 Overview

The goal of our evaluation method is two-fold. First, it offers the ability to compare with the existing approaches by using standard collections, metrics and procedures from the IR community. We compare our approach against the Ide dec-hi with TFIDF, therefore, the procedures used for evaluating these techniques must be adhered. Second, our evaluation method assesses the special feature of our user model, which is the use of knowledge learned over time to modify queries. The procedure, therefore, needs to assess the effects of the users' prior knowledge and combination between the users' prior knowledge and knowledge learned from a query or a set of queries.

4.2 Testbeds

We chose CACM and Medline as our testbed collections because they have been widely used for evaluating the effectiveness of some important RF and QE techniques [13, 4, 8]. CACM contains 3204 documents and 64 queries in computer science and engineering (CSE) domain while Medline contains 1033 documents and 30 queries in the medical domain. The characteristics of the CACM and Medline documents used in our evaluation are shown in Tables 1. We evaluated our user model and TFIDF with Ide-dec hi approach over the entire set

of questions from these two collections because we wanted to obtain a baseline performance for our approach on these two collections.

<i>Attributes</i>	CACM	Medline
Total vectors	3204	1033
Mean length of vector	19.57	53.36
Standard deviation of length of vector	21.91	24.83
Mean frequency of term in a vector	1.61	1.46
Percentage of term with frequency 1	89%	74.78%

Table 1. Characteristics of CACM and MEDLINE documents

4.3 Procedures

Standard procedure applied to Ide dec-hi/TFIDF and user modeling

We follow the procedure laid out by Salton and Buckley [13]. For the Ide dec-hi/TFIDF, each query in the testbeds is converted to a query vector. The query vector is compared against each document vector in the collection. For our approach, we construct a QG for each query in the testbeds in the same way that we construct DGs, in which Link Parser [19] is used. Link Parser sometimes produces incorrect parse trees for the sentences with words which are not found in its dictionary. Therefore, we manually created 27 QGs out of 30 queries for Medline and 21 QGs out of 64 queries for CACM. Medline contains many specialized terms used in the medical domain and CACM contains many specialized terms used in the CSE domain which are not found in Link Parser’s dictionary. We would like to make sure that we have correct QGs to work with. There is no intervention made during the construction of DGs. The main difference between vector representation of TFIDF and our graph representation described above is that the former focuses on frequency of an individual word while ours focuses on the relationship among terms. After we issue each query, the relevant documents found in the first 15 returned documents are used to modify the original query. For the Ide dec-hi/TFIDF, the weight of each word in the original query is modified from its weights in relevant documents and the first non-relevant document. The words with the highest weights from relevant documents are also added to the original query. For our user modeling approach, we start with an empty user model and add the concept and relation nodes to the original QG based on the procedure described in Section 3. We then run each system again with the modified query. We refer to the first run as *initial run* and the second run as *feedback run*. For each query, we compute average precision at three point fixed recall (0.25, 0.5 and 0.75).

Special procedure for user modeling approach

The special procedure assesses the effects of prior knowledge and the combination of prior knowledge with knowledge learned from a query or a group of queries. These requirements lead to our decision to perform 4 experiments:

Experiment 1: We start with an empty user model. We update the initial user model based on relevance feedback and we do not reset our user model unlike the standard procedure above. The user model obtained at the end of this experiment is used as the seed user model for the next 3 experiments.

Experiment 2: We start with the seed user model. For each query, we don't update our user model. This experiment assesses how the prior knowledge helped improve retrieval performance.

Experiment 3: We start with the seed user model and run our system following the standard procedure described above. However, after each query, we reset our user model to the seed user model. This experiment assesses the effects of the combination of prior knowledge and knowledge learned from a given query on retrieval performance.

Experiment 4: We start with the seed user model. For each query, we update our user model based on relevance feedback and we do not reset our user model. This experiment assesses the effects of combination of prior knowledge, and knowledge learned immediately from each query and knowledge learned from previous queries on retrieval performance.

In this procedure, we use the prior knowledge, which is dynamically constructed after Experiment 1 as opposed to using no prior knowledge as in the standard procedure above.

5 Results and Discussions

5.1 Results for standard procedure

The average precision at three point fixed recall of the initial run and feedback run using residual collection of the experiments in standard procedure for CACM and Medline is reported in Table 2. Those in previous publications achieved a slightly better results compared to ours because (i) we used the entire set of queries, while others, for example [4] used a subset of queries; and (ii) we treat the terms from title, author and content equally. Table 2 shows that we achieved competitive performance in both runs for residual and original collections.

The results of our special procedure on user modeling approach are shown in Table 3. Experiment 2 shows that by using the seed user model as prior knowledge for a user, the precision has been increased for the initial runs. Experiments 1, 3 and 4 show that by using our user model, the precision of the feedback runs is always higher using residual and original collections than those of the initial runs. For both collections, we can see that among the four experiments, Experiment 4 performs competitively compared to Ide dec-hi in the feedback run while it offers the advantages of having higher precision in the initial run compared to TFIDF. This shows that we have already retrieved quality documents earlier in the retrieval process than the other approach, leaving less relevant documents

	TFIDF/Ide dec-hi		User modeling	
	Residual	Original	Reisidual	Original
CACM				
Initial run	0.065	0.091	0.067	0.095
Feedback run	0.12	0.2	0.090	0.223
MEDLINE				
Initial run	0.19	0.39	0.212	0.4
Feedback run	0.32	0.54	0.328	0.583

Table 2. Average precision at three point fixed recall for standard procedure

	CACM		Medline	
Experiments	Residual	Original	Residual	Original
Exp 1.				
Initial run	0.073	0.095	0.212	0.446
Feedback run	0.091	0.223	0.344	0.614
Exp 2.				
Initial run	0.075	0.095	0.249	0.512
Exp 3.				
Initial run	0.075	0.095	0.249	0.512
Feedback run	0.11	0.21	0.343	0.609
Exp 4.				
Initial run	0.082	0.095	0.258	0.525
Feedback run	0.11	0.23	0.360	0.625

Table 3. Average precision at three point fixed recall for special procedure

for us to retrieve in the feedback run. The average precisions of experiments 3 and 4 (in which seed user models are used) are higher than those of experiments 1 and experiments in standard procedure for both collections most of the time.

5.2 Discussion

The standard procedure offers us a chance to compare with the TFIDF and Ide dec-hi approaches using their evaluation procedures on the same collections. These queries are as complicated as the ones asked by any real user. However, the evaluation procedures is lightweight and they can be easily used to evaluate adaptive systems before hiring the real subjects. This maintains objectivity and serves as a baseline comparison for future extensions. The special procedure evaluates the long-term effects of knowledge learned in three ways: (i) using the seed user model as prior knowledge, (ii) using the seed user model and updating it with knowledge learned from a query only, and (iii) using the seed user model and updating it with knowledge learned again from a set of queries.

The results show that the retrieval performance increased with all three of these methods. This methodology shows the best performance using a combination of prior knowledge and knowledge learned from a group of queries. For example, in Experiment 4 of the special procedure on Medline, question 7 in the initial run has an added relation "*radioisotop scan - isa - scan*" by the user model and thus has retrieved two more relevant documents in the top 15 than it did in Experiments 1, 2 and 3 (6 relevant documents in the top 15 in Experiment 4 vs 4 relevant documents in top 15 in Experiments 1,2, and 3). We have also applied this method to another collection CRANFIELD [11], and show that our user modeling approach has the potential to improve efficiency, learnability, and interactivity between a user and an IR system by retrieving more highly relevant documents, quickly. Our work here demonstrates this evaluation methodology can be used to assess the impact of knowledge captured by our user model over time to IR process.

6 Conclusion

In this paper, we have reported our evaluation method to assess the effectiveness of our user model with regards to retrieval performance using CACM and Medline collections. The results of this evaluation show how we can compare the user modeling approaches using procedures, collections and metrics of the IR community while still being able to assess special features of the models.

There are issues that we wish to address from this research. Our user modeling approach works best if a user has demonstrated his/her searching styles. So, we will consider re-ordering the queries to effect different search styles (e.g users explore a topic, its subtopics, and then change to a new topic). It will help closely relate the experiment to real life situations while maintaining its objectivity. In this current evaluation, we used the seed user model obtained from Experiment 1. In the future, the seed user model can be created manually (which is likely to achieve even better results) or can be learned from a training query set.

We would like to combine the results of this phase with two other phases [10] to provide a big picture analysis of the overall effectiveness of our user model. This evaluation experiment plays a very important role in the analysis of the overall effectiveness of our user model in terms of improving retrieval and user performance. This data gives us the relevant documents identified by experts who created these collections while the data from our assessments of user performance will give us the relevant documents identified by real users with varying levels of expertise. We will then be able to draw a clear connection between objective and subjective relevancy and how they affect the retrieval performance as well as user performance.

References

1. Balabanovic, M.: Exploring versus exploiting when learning user models for text recommendation. *User Modeling and User-Adapted Interaction* 8 (1998) 71-102

2. Billsus, D., Pazzani, M.J.: User modeling for adaptive news access. *Journal of User Modeling and User-Adapted Interaction* **10** (2000) 147–180
3. Bueno, D., David, A.A.: Metiore: A personalized information retrieval system. In: Bauer, M., Vassileva, J. and Gmytrasiewicz, P. (Eds.). *User Modeling: Proceedings of the Eight International Conference, UM2001, Berlin, Springer* (2001) 168–177
4. Loper-Pujalte, C., Guerrero-Bote, V., Moya-Anegón, F.D.: Genetic algorithms in relevance feedback: a second test and new contributions. *Information Processing and Management* **39**(5) (2003) 669–697
5. Chin, D. Evaluating the effectiveness of user models by experiments. Tutorial at User Modeling conference, Johnstown, Pittsburgh. (2003).
6. Frake, W.B., Baeza-Yates, R.: *Information Retrieval: Data Structures and Algorithms*. Prentice Hall PTR, Upper Saddle River, NJ 07458 (1992)
7. Jensen, F.V.: *An Introduction to Bayesian Networks*. Univ. College London Press, London (1996)
8. Drucker, H., Shahraray, B., Gibbon, C.: Support vector machines: relevance feedback and information retrieval. *Information Processing and Management* **38**(3) (2002) 305–323
9. Magnini, B., Strapparava, C.: Improving user modeling with content-based techniques. In: In Bauer, M., Vassileva, J. and Gmytrasiewicz, P. (Eds.). *User Modeling: Proceedings of the Eighth International Conference, UM2001, Berlin, Springer* (2001) 74–83
10. Nguyen, H.: Capture user intent for information retrieval. In: *Doctoral Consortium at AAAI 2004*. (2004) To appear.
11. Nguyen, H., Santos E. Jr., Zhao, Q. and Wang, H. Capturing User Intent for Information Retrieval. In: *the Proceedings of the 48th Annual meeting for the Human Factors and Ergonomics Society (HFES-04)*. October 2004, New Orleans. (2004). To appear.
12. Salton, G., McGill, M.: *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company (1983)
13. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science* **41**(4) (1990) 288–297
14. Santos, E., Nguyen, H., Brown, S.M.: Kavanah: An active user interface information retrieval application. In: *Proceedings of 2nd Asia-Pacific Conference on Intelligent Agent Technology*. (2001) 412–423
15. Santos, E. Jr., Nguyen, H., Zhao, Q., Hua, W.: User modelling for intent prediction in information analysis. In: *Proceedings of the 47th Annual Meeting for the Human Factors and Ergonomics Society (HFES-03)*. (2003) 1034–1038
16. Santos, E. Jr., Nguyen, H., Zhao, Q., Pukinskis, E.: Empirical evaluation of adaptive user modeling in a medical information retrieval application. In: *Proceedings of the ninth User Modeling Conference*. (2003) 292–296 Johnstown. Pennsylvania.
17. Saracevic T., Spink A., Wu W. Users and Intermediaries in Information Retrieval: What Are They Talking About? *Proceedings of the 6th International Conference in User Modeling* (1997) 43–54 Springer-Verlag Inc.
18. Spink A., and Losee R. M. Feedback in information retrieval. Williams, M., ed., *Annual Review of Information Science and Technology* **31** (1996) 33–78
19. Sleator, D.D., Temperley, D.: Parsing english with a link grammar. In: *Proceedings of the Third International Workshop on Parsing Technologies*. (1993) 277–292
20. Ruthven, I., Lalmas, M.: A survey on the use of relevance feedback for information access systems. *Knowledge Engineering Review* **18** (2003)

An Example of Evaluation applied to a course adapted to learning styles of CHAEA's test

M^a del Puerto Paule Ruiz, Juan Ramón Pérez Pérez, Martín González Rodríguez

HCI Group Research
Dpto. Computer Science of University of Oviedo
C/Calvo Sotelo s/n. 33007 Oviedo
{paule, jrpp, martin}@uniovi.es

Abstract: This paper shows the results of an evaluation of a course adapted to learning styles of CHAEA's test. It is a comparative analysis between an adapted course and a course without adaptation also.

1 Introduction

Normally, the courses published in the Web are thought to get the learning of the students that visit the web site, but the majority of these courses do not include adaptation, which means that the student has to adapt himself to the Course and not the Course to the user. When the user has to adapt to the Course, often the result is not as desired, since the user is not comfortable and will probably not visit the Web site again and the initial goals of learning and diffusion are not carried out.

The adaptation is necessary, but the question is: What type of adaptation: Lexical [1], syntactic [4] or conceptual? The conceptual level seems the most appropriate [7], because this type includes cognitive parameters. These cognitive parameters are very important in the design of the on-line courses.

2 CHAEA's Test

There are a lot of cognitive parameters, but, one of the most important is the learning style. The learning styles are the way of thinking, the way of processing the information, and the way of learning each individual student has. There are a lot of classifications of the learning styles, but the classification selected in this paper is the CHAEA's test.

This test offers an acceptable reliability and validity that has been proved in Spanish Universities [2] and returns the preferences of the student at the time of learning. There are four styles in this classification: Theoretician, Activist, Reflexive and Pragmatic. The test returns a value between 0 and 20 for each style. With these values we get know the learning style. For example, if the student gets 20 in Activist style, he/she is Activist and she/he is going to learn like an Active learner.

Each style has its own characteristics and particularities that are clearly defined by the author [2] but we are going to explain some characteristics that we consider important in order to understand the adaptation [3].

Firstly, the adapted course has been implemented for the Theoretician and Activist learning styles. The selection of these learning styles is because the Theoretician and Activist are the base of Pragmatic and Reflexive learning styles.

Secondly, the Theoretician student likes the theory and she/he doesn't learn with examples or exercises. She/he learns in an inductive way and the contents have to be organized by concepts; however the Activist students like very little the theory and they prefer the exercises. They need a continuous feed-back.

3 Evaluation

In this experiment a Course about HTML is designed [5]. This course is very basic and it consists of six lessons: First Page with HTML, Headings, Paragraphs, and Design with Style, List and Links. It has been adapted for the Theoretician and Activist Users.

3.1 Formulation of hypothesis:

There are two questions:

1. Is the learning with adaptation better than the learning without adaptation?
2. Does the way of evaluating depend on the learning style?, if, do theoretician students resolve only theory questions? Do activist students resolve only active questions?

The experiment has to prove the following hypothesis:

1. The learning with the Adaptation Model is more effective than the learning without the Adaptation Model.
2. The evaluation of the knowledge does not depend on the learning style. The most important thing is the learning itself and not how this learning is evaluated.

3.2 Identification of variables

Two variables have been selected to develop the experiment: The learning style (independent variable), and the result of the evaluation of knowledge (dependent variable).

3.3 Population and Sample

The population consists of students of the subject Information Systems of the Business School of University in Oviedo.

The sample is composed of 54 students. 27 of them are theoreticians and 27 of them are activists. For all of them it is the first time that they have been registered in the subject and they do not know anything about HTML.

There are three groups: A Control group, an experimental group and the Non_Adapted_Test Group. These groups are homogeneous in their composition, so, there are nine theoretician and nine activist students in each group. The distribution of students is random in each group.¹

3.4 The design of th experiment

There are two types of courses:

1. An adapted course: This course has different interfaces depending on the student [6]:
 - The Theoreticians have more theory than exercises and the theory is organized in concepts.
 - The Activists have more exercises and more examples than theory.
2. A course without adaptation: This course offers the same interface with theory, examples and exercises for each lesson.

There are also two types of tests:

1. Test with questions adapted to learning styles
2. Test with questions without specific adaptation, so, there are questions of both learning styles.

The experimental sessions are developed in the Business School of University of Oviedo. Each session is 110 minutes long.

These sessions have four parts:

1. Attitude Test².
2. Chaea's Test.
3. Surfing on the Course
4. Acquired knowledge test.

The 1st and 2st parts are equal for all groups, but the 3st and 4st parts are different depending on the group.

The control group students have to surf the course without adaptation and they have to do a test adapted to the learning style.

The experimental group students have to surf the adapted course and they have to do a test adapted to the learning style.

The Non_Adapted_test group have to surf the adapted course and they have to do a test without adaptation.

The students of these groups are different.

The test is formed by 15 multiple answer questions. At the end of it, the students know the score.

The test has 3 questions for each lesson of the course.

¹ Each group has majority of girls than boys because more girls than boys registered in the subject Information Systems.

² This test is in bases om Likert Scale with numeric answers where 1 represents the lowest agreement and 5 represents the highest agreement.

Each correct answer scores 1 point and each incorrect answer scores -1 point and the non answered question scores 0 points.

At the beginning of the session the student is told that it is part of a teaching quality evaluation of the University of Oviedo, so the student is not conditioned by the test.

3. 5 Results

With this experiment, information is about the attitude and what knowledge the students acquire after surfing the course.

With this information, it is possible to determinate if the adaptation has an influence on the learning and if the way of evaluating depends on the learning style.

The statistical software SPSS is used to get this analysis.

Firstly, it is necessary to determine if the groups had a similar attitude: A Tstudent is applied to prove this³.

Secondly, it is necessary to check if the results of the control group are better than the results of the experimental group.

Thirdly, we must compare the results of the experimental group (adapted test) and the Non_Adapted_test group (test with a mixture of theoretician and activist questions).

The results of this analysis are shown in the appendix.

In the first part of this analysis, all groups (Control Group, Experimental Group and Non_Adapted_Test group) have a normal distribution for each learning style. Besides, the attitude has a normal distribution in Control Group and Experimental Group.

In the second part of this analysis, the control Group and the Experimental Group have a homogeneous variance and, the Experimental Group and the Non_Adapted_Group have a homogeneous variance too. The Attitude variance is homogeneous in the Control Group and Experimental Group. So, it is possible to apply the TStudent to get the differences and the improvements.

The application of TStudent in the Control Group and Experimental Group returns an improvement, so, the scores obtained in the Experimental Group are slightly better than the scores obtained in the Control Group. This result proves the first hypothesis: There is an increment of learning with the adaptation.

The application of Tstudent in the Experimental Group and Non_Adapted_test returns that the difference is not significant, so, the evaluation is not important, the most important thing is the learning. When a student learns, she/he can answer any type of questions. The evaluation is independent of learning style.

³ TStudent compares independent samples (for example, the different groups scores) and it is used to know if the differences are significant. The TStudent's proof requires normal distributions to apply it. The Shapiro-Wilk test is used to get this, it is also necessary that the variances are homogenous; the Levene's Proof is used to do this verification.

4 Conclusions

The evaluation offers a determinant conclusion: adaptation of contents versus adaptation in the evaluation of contents. The most important thing is the adaptation of contents, and not the adaptation in evaluation of the same ones.

Learning increases with the adaptation of contents, but there are not differences between the evaluations of adapted or not adapted contents. The process of teaching-learning has to focus on the adaptation of contents and not on the evaluation of the contents.

A course that respects the learning style makes the learning more efficient. Internet is the most adequate way to do this adaptation. The contents can be shown in a personal way to every user of the net: this is the power of the Web, since the same contents it can come in different ways to different individuals and this is a fundamental idea : The same knowledges explained in a different way depending on the characteristics and particularities of every user.

Acknowledgements. This research is supported by University of Oviedo, Proyecto MB-04-534-4.

Appendix

Note: Significance level of 95% (0, 05)

1. Normal Distribution for Control Group in Final Score (Shapiro-Wilk)

Learning Style	W	Sig
Activists	0,946	0,637>>0,05→ Distribution can be normal
Theoreticians	0,871	0,163>>0,05→ Distribution can be normal

2. Normal Distribution for Experimental Group in Final Score (Shapiro-Wilk)

Learning Style	W	Sig
Activists	0,854	0,092>>0,05→ Distribution can be normal
Theoreticians	0,854	0,091>>0,05→ Distribution can be normal

3. Normal Distribution for Not Adapted Group in Final Score (Shapiro-Wilk)

Learning Style	W	Sig
Activists	0,828	0,064>>0,05→ Distribution can be normal
Theoreticians	0,828	0,064>>0,05→ Distribution can be normal

4. Normal Distribution of Attitude in the Control Group in Final Score (Shapiro-Wilk)

Learning Style	W	Sig
Attitude	0,924	0,223>>0,05→ Distribution can be normal

5. Normal Distribution of Attitude in the Experimental Group in Final Score (Shapiro-Wilk)

Learning Style	W	Sig
Attitude	0,924	0,223>>0,05→ Distribution can be normal

6. Homogeneity of variance between the scores of Control Group and Experimental Group for each learning style

Learning Style	W	Sig
Activists	3,115	0,097>>0,05→ Variances are homogenous
Theoreticians	4,283	0,006>>0,05→ Variances are homogenous

7. Homogeneity of variance between the scores of Experimental Group and Not Adapted Test Group for each learning style

Learning Style	W	Sig
Activists	0,142	0,711>>0,05→ Variances are homogenous
Theoreticians	0,139	0,715>>0,05→ Variances are homogenous

8. Homogeneity of variance between the scores in the Likert Scale for the attitude

Levene	Sig
0,002	0,899>>0,05→ Variances are homogenous

9. TStudent for Control Group and Experimental Group

Learning Style	gl	Sig
Activists	16	0,001<<0,05→ The significant improvement
Theoreticians	16	0,002<<0,05→ The significant improvement

10. TStudent for Experimental Group and Not Adapted Test Group

Learning Style	gl	Sig
Activists	16	0,567>0,05→ The different is not significant
Theoreticians	15	0,567>0,05→ The different is not significant

References

1. Abascal J., Arrue M., Fajardo I., Garay N., Tomás J. Use of Guidelines to automatically verify web accessibility. Universal Access in the Information Society.

Special Issue on "Guidelines, standards, methods and processes for software accessibility" (Guest editors: Gulliksen J., Harker S., Vanderheiden G.). Accepted for publication.

2. Alonso, C.M., Gallego, D.J., Honey, P. "Los Estilos de Aprendizaje". 4ª Edición. Ediciones Mensajero.
3. Del Moral Pérez and et al. "Rur@lnet: Un espacio para la teleformación y el trabajo colaborativo". In 2nd European Congress on Information Technologies in Education and Citizenship: a critical insight, 2002.
4. Gilbert, J.E., Han, C. Y. "Adpating instruction in search of 'a significant difference'". Journal of Network and Computer Applications, 22.
5. Gross Begoña. "Diseños y programas educativos". ISBN: 84-344-2604-8, 2002
6. Paule Ruiz, Mª Del Puerto, Ocio Barriales, Sergio, Pérez Pérez, Juan Ramón, González Rodríguez, Martín. Feijoo.net. "An Approach to Personalized E-learning Using Learning Styles". Web Engineering, International Conference, ICWE03. Lecture Notes in Computer Science 2722. ISBN: 3-540-40522-4 Springer Verlag Berlin . ISSN: 0302-9743, 2003
7. González Rodríguez, Martín, López Pérez, Benjamin, Páule Ruíz, María Del Puerto, Pérez Pérez, Juan Ramón. "Dynamic generation of Interactive Dialogs Based on Intelligent Agents". Lecture Notes in Computing Science Vol. 2347, Springer, May 2002. Pp. 564 - 567. ISSN 0302-9743, 2002

Evaluating a General-Purpose Adaptive Hypertext System

Chris Staff

University of Malta,
Department of Computer Science and AI,
Msida MSD 06, Malta

Abstract. We describe the evaluation process of HyperContext, a framework for general-purpose adaptive and adaptable hypertext. In particular, we are interested in users' short-term, transient, interests. We cannot make any prior assumptions about a user's interest or goal, as we do not have any prior knowledge of the user. We conducted evaluations on two aspects of HyperContext. One evaluation was completely automated, and the other involved participants. However, the availability of a test collection with value judgements would be a considerable asset for the independent and automated evaluation of adaptive hypertext systems in terms of cost, reliability of results, and repeatability of experiments.

1 Introduction

HyperContext is a framework for adaptive and adaptable hypertext [8], [9]. We are currently using the HyperContext framework as part of the University of Malta's contribution to the Reasoning on the Web with Rules and Semantics (REWERSE) FP6 Network of Excellence¹.

HyperContext focuses on building and maintaining a short-term user model to provide adaptive navigation support. We begin a user session with an empty user model and we add to the model as a user navigates through hyperspace and interacts with the system.

A proof of concept HyperContext application has been evaluated. We had devised an evaluation strategy for HyperContext in 1999. However, due to a number of reasons, including hardware failure, the original evaluation strategy was abandoned. We eventually settled on a partially automated approach that did involve some participants, but which was less reliant on human participants.

We are satisfied that the results of the automated evaluation show that the adaptive features of HyperContext can guide users to relevant information. We feel that our automated evaluation benefited from the fact that HyperContext assumes an initially empty user model that is then populated during short interactions with the system. Part of the evaluation involved showing users a series of documents (representing a short path through hyperspace) followed by two other documents in a random sequence. One of the two documents was recommended

¹ staff.um.edu.mt/mmon1/research/REWERSE/

by HyperContext using a user model that would have been generated had the user actually followed the path through the first 5 documents in a HyperContext hyperspace. The other document was also a recommended document, but the user model used to make the recommendation was derived in a different way. We are able to demonstrate that the second recommendation is based on a user model built on a Web-based, rather than a HyperContext-based, hyperspace. The evaluation is similar to an approach using with- and without-adaptive functionality [6], but we show that the without-adaptive functionality system is equivalent to the World Wide Web. The results of the evaluation are reported extensively in [8] and [9]. In this paper we concentrate on reporting the evaluation *process* and our opinion on its suitability for the evaluation of adaptive hypertext systems.

2 Objectives of the Evaluation

Before we discuss HyperContext and the evaluation strategies, we present our motivation and objectives for evaluating HyperContext. An adaptive hypertext system may use adaptive navigation techniques to guide users to relevant information in hyperspace [2]. As HyperContext utilises adaptive navigation techniques almost exclusively (there is limited support for adaptive presentation, but this was not the focus of our research), we expected that a HyperContext user would find relevant information faster than a user using a non-adaptive equivalent, as HyperContext would recommend links and paths to users, assuming that the user model accurately reflected the user's needs and requirements.

HyperContext is a general-purpose system for use in a heterogeneous information space, such as the WWW. Consequently, unlike an educational hypertext system, we cannot make certain assumptions about our users. For instance, the goal of a user of an educational system is likely to be to increase his or her knowledge of the subject contained within the system. As the domain is restricted, it is possible to pre-test or "interview" the user to initialise the user model with useful information. The users of a general-purpose hypertext system that focuses on collection of short-term information are not so helpful. A short-term user model is likely to be at its most useful when the user is navigating through territory with which he or she is unfamiliar and when the user's interest in the information is significant but transient. For instance, a user may have some task to perform and some information is required to perform that task. Although the completion of the task is dependent on obtaining the information, the user's interest in it lasts only as long as it takes to complete that aspect of the task. What motivates us is the challenge of recommending useful links (i.e., links that are likely to lead the user to relevant information) when we initially know little or nothing about the user's interests, goals, and expertise. However, motivating evaluation participants to the degree that they will search for information that they know little about but really need under evaluation conditions is hard. Either the prototype software under evaluation will need to be robust enough to use on the Web at large (in which case participants can use the system in their

own time), or a smaller Web space will be converted for use with the hypertext system (so that HTML pages, for instance, will be free from error), in which case the chances of finding adequately motivated participants is greatly reduced.

For our evaluation, we converted part of the World Wide Web Consortium's (W3C) website² to a HyperContext hyperspace. We chose the W3C site because it is about Web standards ranging from HTML to Web-HCI issues, so we reasoned that the site was designed to be easy to use, consistent, and relatively free from (HTML) errors, which would ease processing. An explanation of what is involved in the conversion is given in section 3. We also show in subsection 4.1 that without the adaptive features provided by HyperContext, the converted site is equivalent to the original Web site.

3 Generating a HyperContext hyperspace

In a hypertext, the same document can be the destination of many different links. Consequently, the same document may be reached along different paths. It is possible that users who reach the same document following different paths may be looking for different information, or may have reached the same document for different reasons. Such users are likely to interpret the information in the document differently, depending on the other documents in this session the user has so far read and any other knowledge and interests that the user might have. If we are to individualise link and path recommendation knowing only the user's path of traversal through hyperspace, then we need to understand how the information in the child (destination) document is related or relevant to the information in each of the child's parents.

On the Web, web pages range from short and single topic to huge, multi-topic documents. The length of a web page is not a good indicator of the number of topics it is likely to contain. Should information about all topics in a document be added to the short-term user model, in the hope that eventually the dominant topic will float to the surface? Should we use topic distillation algorithms to split up a document into its different topics, and compare each topic to the topic of the region in the parent that the user followed to reach this child? We opted for the second approach to determine the relevant terms in a document visited by the user. A document *interpretation* is a vector of term weights which partially describes a document in the context of a parent. A document has at most $n+1$ interpretations: one for each of its n parents, and an additional one (the *context-free interpretation*), that does not decompose a document into its different topics, which is invoked if a document is accessed directly rather than by following a link to it. To convert the W3C web site to a HyperContext hyperspace we created interpretations for each (HTML) document. A link in the new adaptive hyperspace is retained if the topic distillation algorithm determines that there is sufficient similarity between the topics in the source and destination documents. The user model is updated each time the user traverses a link, using information derived from the visited document's interpretation.

² www.w3.org

3.1 The User Model

The short-term user model is based on the interpretations of documents that the user has accessed during the current session. The user model is used to recommend links each time a document is accessed. A query may also be extracted from the user model and submitted to an information retrieval system to retrieve relevant interpretations if these have been previously indexed.

3.2 Evaluating HyperContext

As we discussed in section 2, our goal is to direct users to relevant information faster than they would be able to find it themselves, particularly when they are unfamiliar with the topic. We describe our original evaluation strategy in section 4. In section 5, we describe the actual strategy we used to evaluate HyperContext. In this paper, we concentrate on the evaluation *process*. The evaluation results are discussed in detail elsewhere [8], [9].

4 Evaluation Strategy 1

The empirical study that we had originally planned was to involve three groups of six participants each. Of the 18 participants, 6 each were previously judged to be novice, intermediate, and advanced information seekers. The initial study involved 36 participants who were set 15 general knowledge information seeking tasks. They were allowed to use any information source (search engine, web directory, their own memory) they liked, but had to indicate if they already knew the answer. For each task, the student had to write down a URL containing the answer (or URLs, if the answer spanned a number of web pages). The information seeking tasks were pre-tested to ensure that the answers were available on the Web.

Each participant's performance for each task was compared to the average time to perform each task (from among those participants who did not already know the answer). Participants who generally arrived at a solution faster than average were considered advanced information seekers, those who were generally much slower at finding information were considered to be novice, and the others were considered intermediate. 6 people were to be randomly selected from each group to participate in the HyperContext evaluation.

A HyperContext Evaluation group was to consist of two novice, two intermediate, and two advanced information seekers. Each group would have an identical set of tasks to perform. The tasks were designed to find technical information, rather than general knowledge as used in the experiment to classify participants. One group would act as the control group, the second and third groups would both use a HyperContext-enabled version of the W3C web site, but the algorithm used to construct the user model would be different. Once again, the performance of the two HyperContext-enabled groups would be compared to the performance of the control group, where we can show that the control group would have used

the equivalent of the W3C web site. Each group would have access to the same information search and retrieval system. The control group would have access to an index generated from the original, unmodified documents, whereas the other groups would have access to an index that also contained an index of document interpretations (document interpretations are discussed in section 3).

4.1 Is a without-adaptation HyperContext equivalent to the Web?

The HyperContext hyperspace created from the W3C web site for use in the evaluation (section 3) can be considered equivalent to the original W3C web site if adaptivity is disabled. By default, the context-free interpretation of a document consists of a vector of term weights for all terms that occur in the document, rather than just those terms that are considered relevant to the parent, when a link is followed. In the disabled version of HyperContext, all link traversals invoke the context-free interpretation, so the interpretation of the document is the same regardless of how the document is accessed. This behaviour is equivalent to the behaviour on the Web. Regardless of how any page is accessed, normally there is absolutely no difference in or about the page that was accessed.

5 Evaluation Strategy 2

Due to a number of unfortunate incidents, including hardware failure resulting in total data loss, and looming deadlines, the intended strategy outlined in section 4 never progressed beyond the first stage of classifying participants as novice, intermediate, and advanced information seekers. By the time the HyperContext hypertext and related data were recovered, there was simply not enough time to re-run the original classification of participants (because their information seeking skills were bound to have changed over time [3]), conduct the rest of the evaluation and analyse the results. Instead, we decided to separate the evaluation of some of the functionality from the evaluation requiring user participation [5]. We developed one completely automated experiment to test our hypothesis about the improved ability to locate relevant information in a HyperContext hyperspace. A second experiment required anonymous Web-based participation from users to judge whether documents recommended by HyperContext were relevant to information they had read on a pre-determined path through a HyperContext hyperspace.

5.1 Locating relevant information

The number of links on a page, coupled with the lack of a link semantics in HTML increases cognitive overhead. A user must decide whether or not to follow a link. Adaptive educational hypertext systems may make use of visual link adaptation to indicate that a link may be followed with profit, or should not yet be followed, e.g., [11]. Alternatively, forms of link hiding [2] may be used, in which users are discouraged from following links unlikely to lead to relevant information. In

either case, this is a form of hypertext partitioning - separating the non-relevant parts of hyperspace from the relevant.

In HyperContext, a user visiting a document actually visits an interpretation of that document. In Section 3 we explained that an interpretation is a vector of term weights, and that different interpretations of the same document may have different vectors of term weights. For instance, in one such vector, some term t_n may have weight w_x . In another interpretation of the same document, the same term may have the same weight, or a completely different weight, depending on how significant the term is to the context of the topic of that document's parent. Interpretations are slightly more complex, however. One interpretation of a document may have link anchors which may or may not be active in other interpretations (a form of link hiding). Additionally, even if the same link is active in several interpretations of the same document, the destination of the link may change depending on the interpretation (figure 1). In this way we are able to partition a HyperContext hyperspace, potentially separating the non-relevant from the relevant.

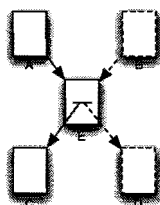


Fig. 1. Link in doc E leads to C if entered from A, and to D if entered from B.

To determine if multiple interpretations of information can adequately partition a hyperspace so that a user can be led to relevant information, we count the number of nodes that must be visited starting from some arbitrary start node until we reach a relevant node. A relevant node is just some randomly selected node that is at least 2 link traversals away from the start node. We compared two adaptive solutions to two non-adaptive solutions, measuring overall performance and the performance of each approach as the path length grew. The adaptive solutions were based on a HyperContext enabled converted W3C hyperspace, and the non-adaptive solutions were based on the original W3C web site. The premise is that the optimal solution is one that finds the shortest path between the start node and the target, relevant, node. The least optimal solution is likely to be based on a breadth-first or depth-first brute force search (depending on the "shape" of the hypertext graph), essentially following the links in the order of least likely to lead to the target node. For this experiment we traversed the links in the order they occurred in a document, using a hybrid approach. We process nodes breadth-first until we encounter the target node. We then prune the graph

of accessed nodes, eliminating all visited nodes to the right of the shortest path between the start node and the target node (figure 2). This is the equivalent of a depth-first search guarded by the known depth of the target node. If the best link to follow always happens to be the first link in a document, then this approach will give results similar to the optimal solution. However, unless the best link is always the last one in a document, then this approach will give better results than the least optimal solution, because nodes which did not need to be visited will be not be counted. This approach yielded paths of maximum length 5 (four link traversals from the root).

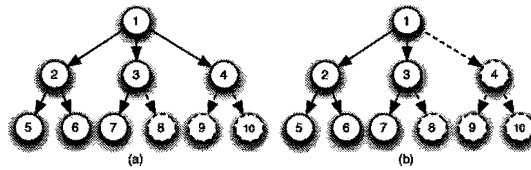


Fig. 2. Node 7 is the target node. (a) Solid nodes are visited in breadth-first search; (b) hybrid depth-first marks node 4 as unvisited.

An algorithm that partitions the hyperspace may decrease the span of the graph, and hence improves the speed with which a target node can be reached, even when a brute-force approach is taken. The efficiency may be decreased if a relevant node is made either unreachable or reachable by a longer traversal of the graph if the hyperspace is partitioned badly (figure 3). In either case (adaptive or non-adaptive) the efficiency may be further improved by introducing a link ordering algorithm that ranks links in a document according to the likelihood that they will lead to the target node. The link ordering algorithm compares the current node's children (a lookahead of 1) to the target node. Links in the current document are traversed in the order of degree of similarity between the link destination and the target node. In the experiments with the adaptive version of the hypertext, the interpretation of each child (section 3) is used by the algorithm, rather than the context-free interpretation of the child used in the non-adaptive version.

5.2 Evaluating Document Recommendation

In the second part of the evaluation, we prepared a number of paths through hyperspace that all involved exactly four link traversals (for consistency with the maximum path length reported in subsection 5.1) through different documents. If a document was re-visited on a path, the path was not selected for the experiment. Two user models were maintained. We assumed that the first document on the path was the root of the path, and that both user models were empty at this point. Each user model was updated following a link traversal to

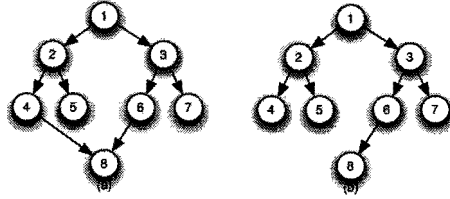


Fig. 3. Partitioned hypertext: (a) before, and (b) after. Node 8 is no longer reachable from node 4 in (b) and so may take longer to reach.

the next document on the path. On reaching the last (fifth) document on the path, two queries were generated, one from each user model, and submitted to our search engine. The first document recommended by each user model was noted. Eventually, participants were asked to give relevance judgements about each recommended document having first read all documents on the path.

A term weight vector based on the interpretation of each visited document on a path is used to update the first user model ($UM_{adaptive}$). For the second user model ($UM_{control}$), the context-free interpretation of the document is used³. If both user models recommended the same document, the path was considered inapplicable for evaluation purposes and was discarded. Eleven conforming paths of length five were randomly selected. The documents on the path and the documents recommended by each user model were placed on-line and hosted by a Web server for 25 days. Members of staff in the Department of Computer Science and AI at the University of Malta and its student population were invited via e-mail to participate in the on-line evaluation. Participation was totally anonymous and could be carried at the participant's leisure from a location of their choice. Participants were asked to read each of the first five documents in a path in the order they were displayed. They were then shown two recommended documents (one after the other) and asked to give a relevance judgement about each.

We used a 4-scale of relevance judgements (highly relevant, quite relevant, quite non-relevant, highly non-relevant), rather than the two (relevant, not relevant) normally used [10], because we expected both user models to make recommendations of at least slightly relevant documents. Participants were not told the order in which recommended documents would be displayed. They did not know which document was recommended by $UM_{adaptive}$ and which was recommended by $UM_{control}$. The sequence was set randomly.

5.3 Summary of Results

The results of the evaluation are reported extensively in [9] and summarised in [8]. To locate relevant information we measured the difference between the best

³ This is the equivalent of the Web version of the document (section 4.1).

case scenario (the shortest path between two nodes), the worst case (the longest path assuming that we know the level depth of the target node), and the adaptive solutions. The adaptive solutions outperformed the non-adaptive ones as path length increased. If the target node was 3 or 4 link traversals from root, then the adaptive solutions found the target node having visited less intermediate nodes than the non-adaptive approaches. This performance was reversed for target nodes that were up to 2 links traversals away from the root.

For the second part of the evaluation, two user models were used to recommend documents to users using an adaptive and a non-adaptive approach respectively. At face value, documents recommended by the non-adaptive approach were considered more relevant than those recommended by the adaptive approach. However, if time spent reading a document is an indication that a document is skim read or read closely (deep read), then readers tended to consider relevant the document recommend by the adaptive approach when the documents were deep read, and those recommended by the non-adaptive approach if the document was skim read. However, this is an assumption because although we measured the amount of time spent reading each document on a path users were not asked to confirm whether they skim or deep read the documents.

6 Conclusion

One main and significant difference between general-purpose adaptive hypertext systems, like HyperContext, and adaptive educational hypertext systems is that our evaluation participants did not necessarily have any motivation to read about or learn about the information contained in our hyperspace (Web standards). In educational hypertexts, there may be more scope for finding participants who are interested in learning what the system is teaching. We feel that HyperContext would have benefited from evaluation by participants who use it to guide their search for information that they are motivated to obtain. However, setting up such experiments can be complex and expensive [4]. For example, the Alberta Ingenuity Centre for Machine Learning pays an honorarium to Web-based participants in the evaluation of LILAC⁴.

Creating test collections with value judgements for adaptive hypertext systems may make the results of automated evaluation more reliable and comparable, as has been the case with information retrieval and systems for some decades [1]. Perhaps the most common criticism of this approach, and one that could also effect adaptive hypertext systems, and not merely because some, like HyperContext, make use of information retrieval systems to make recommendations, is that *relevance* is highly subjective. The Text Retrieval Conference uses “pooling” to set relevance judgements for documents in test collections [10], [7].

We automated some of the evaluation process for HyperContext. We selected the algorithm for updating the user model, and we used a simple topic distillation algorithm to create interpretations of documents based on each of their

⁴ www.web-ic.com/lilac/honorarium.html

parents to partition hyperspace so that we can more quickly locate a target document presumed to contain relevant information. Of course, this automated experiment alone was insufficient to conclude that users would actually find the recommended documents relevant, so we then invited participants to provide relevance judgements for documents that were recommended after the participants had read 5 documents on a path through a converted W3C web site.

We use a short-term user model that is initially empty to collect information about a user's interests as the user navigates through hyperspace. This is the only way in which the user model can be updated. If a user is not permitted to use a search engine to locate information, or to jump directly to pages using their URL, or to directly edit the user model, but can only follow paths through the information space, then the user model of all users following the same path will be updated in the same way, and the same recommendations will be made. If we can also know in advance which links and documents should be recommended at each stage, then it should be possible to create a test collection with relevance judgements that can then be used for automated evaluation.

References

1. Baeza-Yates, R. and Ribeiro-Neto, B.: *Modern Information Retrieval*, Chapter 3: Retrieval Evaluation (1999) Addison Wesley:US
2. Brusilovsky, P.: *Methods and techniques of adaptive hypermedia*. *Adaptive Hypertext and Hypermedia* (1998) 1–43. Kluwer Academic Publishers:Netherlands.
3. Chin, D. N.: *Empirical Evaluation of User Models and User-Adapted Systems*. *User Modeling and User-Adapted Interaction*, v. 11, n1-2 (2001) 181–194
4. Del Messier, F., and Ricci, F.: *Understanding Recommender Systems: Experimental Evaluation Challenges*. Weibelzahl, S., and Paramythis, A. (eds) *Proceedings of the Second Workshop on Empirical Evaluation of Adaptive Systems in conjunction with UM2003* (2003) 31–40.
5. Herder, E.: *Utility-Based Evaluation of Adaptive Systems*. S. Weibelzahl and A. Paramythis (eds), *Workshop on Empirical Evaluation of Adaptive Systems User Modeling 2003* (2003) 25–30
6. Höök, K.: *Evaluating the utility and usability of an adaptive hypermedia system*. *Proceedings of the 2nd international conference on Intelligent user interfaces*, Orlando, Florida, United States (1997) 179–186
7. Smeaton, A. F., and Harman, D.: *The TREC Experiments and their impact on Europe*. *Journal of Information Systems* (1997)
8. Staff, C. D.: *The HyperContext framework for Adaptive Hypertext*. *Proceedings of the Thirteenth ACM conference on Hypertext and Hypermedia*, College Park, Maryland, USA (2002) 11–20
9. Staff, C. D.: *HyperContext: A Framework for Adaptive and Adaptable Hypertext* (2001) PhD thesis, University of Sussex.
10. Voorhees, E.: *Overview of TREC 2003*. *Proceedings of the Twelfth Text REtrieval Conference*. (2003)
11. Weber, G. and Brusilovsky, P.: *ELM-ART: An adaptive versatile system for Web-based instruction*. *International Journal of Artificial Intelligence in Education* 12 (4), Special Issue on Adaptive and Intelligent Web-based Educational Systems (2001) 351–384.

Evaluation Experiments and Experience from the Perspective of Interactive Information Retrieval

Ross Wilkinson

Mingfang Wu

CSIRO ICT Center,
Melbourne, Australia
{Ross.Wilkinson, Mingfang.Wu}@csiro.au

Abstract. It has long been a tradition of evaluating information retrieval systems with very simple user models and very simple tasks: the task is to retrieve relevant documents to a user need described by a query. TREC, the Text REtrieval Conference sponsored by NIST, raised the bar by providing large scale collections, well defined user needs, independently judged documents, and a specified form of success. Groups from around the world all tackled this same task that allowed wide analysis of just what factors influenced system performance.

Yet there was concern, as system performance improvement did not always lead to human performance improvement, so a concerted effort to study how people interact with information retrieval systems was undertaken in the Interactive Track of TREC. This paper describes this track, some of the experiments that we have undertaken in this track, and highlights some of the real problems in such evaluation.

There are two key issues that we have often observed in interactive information retrieval. The first issue is that human preference is often not correlated with human performance. Consequently, evaluation that relies solely on either form of evaluation is not reliable. The second issue is that genuine improvements are very difficult to validate, as system variation tends to be dominated by task variation and user performance variation. Consequently, the statistical power of these experiments, often already very expensive to conduct because of user participation, can be quite low. Thus we argue for staged experiments where only very "obvious" system performance gains are explored.

In the end, simple performance measures have proved less helpful than deeper analysis of just how people interact with their information systems.

1 From System-oriented Evaluation to User-oriented Evaluation

Information Retrieval (IR) research has long been driven by evaluation. The evaluation was largely on the system part (or so called batch mode evaluation). The basic testing environment is to build a test collection which includes 1) a collection of documents to search on, 2) a set of queries to search for, 3) and the relevance judgement about each document with respect to each query. A testing system is then fed with a set of queries and scored according to its ability to retrieve relevant documents.

In the early stages (of the field), each research groups built their own small test collection (with hundreds of documents and a few queries). As a result, it was hard to compare systems across research groups and generalize the experimental results. The

TREC [7][14] is a major initiative to address these issues. The goal of TREC is to “encourage research in text retrieval based on large test collection”, and to “provide a common task evaluation that allows cross-system comparisons”. Compared to other previous test collections (such as Cranfield test collections [6]), the TREC test collection enlarges greatly the size of document collection, provides more realistic queries, and uses more realistic relevance judgments made by independent assessors. In addition, TREC also provides specialized test collections for different test tasks (tracks), such as: filtering, high precision, question answering and so on.

This kind of evaluation is designed to allow one parameter to be manipulated at a time - this is good for comparison of individual components embedded within systems. The evaluation criteria are usually: precision (the proportion of retrieved documents that are relevant) and recall (the proportion of relevant documents that are retrieved), measuring accuracy and coverage. This evaluation technique enables system testing to tell us what (component and algorithm) works and what doesn't. Such experiments are usually repeatable.

This kind of system-oriented evaluation is essential for creating more effective and efficient systems for retrieving and organising information, although it has its drawbacks. For example, it treats the search as a one off session, the relevance judgement is usually binary and held constant and users are abstracted away from the search loop. In reality, relevance is dynamic, situational, cognitive, and context dependent [5], and users are often involved in an extended search process, especially now that accessing information through the web has become a routine part of life.

Actually since the beginning of the field, there have been active discussions and debates on addressing the issues of the evaluations of IR as an interactive system, of the including users in the evaluation and in the use of IR systems. It has been argued that the evaluation of IR systems includes not only the retrieval effectiveness, but also the utility, satisfaction and use [8] [10][11][12][13]. This is increasingly recognized as we have seen that the effectiveness of a system may not be automatically transferred to the user performance [4] and it is increasingly hard to achieve further improvements in retrieval accuracy.

Since 1995 (TREC4), the TREC started to set up the interactive track. The goal of this track is “to investigate searching as an *interactive* task by examining the *process* as well as the *outcome*” [15]. Like the batch model evaluation, this track is also a common platform for research groups to evaluate their systems, and a forum for researchers in this area to discuss how to evaluate an IR system as an interactive process. Unlike the batch model evaluation, the interactive track includes the user. The user is given a scenario that simulates a real world search problem, and is asked to perform the search and made their own judgement on what information is relevant and what is not. The evaluation criteria display a mixed flavour of IR and HCI, it includes the user performance and accessibility, usability and user satisfaction.

Each year, the track defines a task and develops a set of topics. The topic is a brief statement of the task (or information need). A common test collection is stipulated and a common evaluation criteria and procedure is recommended. The criteria on user performance are usually task dependent, but overall, they are measures of the success a subject in completing a specified task. During the experiment, a set of instruments are used to collect subjects' subjective evaluation of the testing systems. The instru-

ments include an entry questionnaire to collect subjects' demographic information and search experience, a pre-search questionnaire for each topic to get subjects' familiarity with the search topic, a post-search questionnaire for each topic to collect subjects' opinion on the search experience and the perceived task completeness, a post-system questionnaire for each system to collect subjects' use experience of that system, and finally, an exit questionnaire to let subjects compare the two systems using a given set of criteria.

The experiment design concerns three factors: system, topic and subject. The experimental design should be able to measure separately the effect of topics, subjects and systems as well as gather some information about the strength of expected interactions between system and topic, topic and subject, as well as subject and system. Thus a factorial design, Latin-square experiment design, is recommended by the TREC interactive track. Table 1 shows an example of such a design.

As we can see, this kind of design is comparative in nature. Indeed, in the past, the participating groups compared various feedback methods (e.g implicit versus. explicit [2]), answer organizations (e.g. ranked list versus clustering versus visualization [1][9][16]), document summaries (e.g. general summary versus. task-biased summary [17]), the transfer of system performance to the user performance [4] and many more (please refer: <http://trec.nist.gov/pubs.html>).

Table 1. Minimal Latin-square experimental design
(T: test system, C: control system B1 and B2: two blocks of topics)

Subjects	System, Topic	
1	T, B1	C, B2
2	C, B2	T, B1
3	T, B2	C, B1
4	C, B1	T, B2

2 Our Experience

We have been interested in the effect of different delivery methods on the user's performance of a certain task. Through the interactive track platform, we evaluated and compared an organization of clustering search result with a ranked list, and query-biased document summary with generic document summary on the question answering task, and a topic distillation oriented delivery and a general ranked list on the topic distillation task. In this paper, we will just discuss our participation in the question answering task.

The question answering task here is not simply to find a single fact. The task involves collecting multiple independent data items (called instances) from one or more resources and synthesizing them into an answer. Figure 1 shows an example of such a search topic as given to subjects.

The assessment of the user's performance measures the ability of a subject to identify documents that contain topic instances. It is measured by instance recall and

instance precision, here instance recall equates to the proportion of the known topic instances contained in the documents identified by a subject, and instance precision the proportion of the documents identified by a subject that were deemed to contain topic instances. The assessment process, therefore, provides indirect or, more accurately, circumstantial evidence of the effectiveness of the interactive system's ability to help the subject develop an answer to the information need represented by the interactive topic.

<p>Title: Ferry Sinkings</p> <p>Description: Any report of a ferry sinking where 100 or more people lost their lives.</p> <p>Narrative: To be relevant, a document must identify a ferry that has sunk causing the death of 100 or more humans. It must identify the ferry by name of place where the sinking occurred.</p> <p>Detailed of the cause of the sinking would be helpful but are not necessary to be relevant. A reference to a ferry sinking without the number of deaths would not be relevant.</p> <p>Aspects: Please save at least one RELEVANT document that identifies EACH DIFFERENT ferry sinking of the sort described above. If one document discusses several such sinkings, the you need not save other documents that repeat those aspects, since your goal is to identify different sinkings of the sort described above.</p>

Figure 1. An example topic

Investigation 1: Clustering structure versus ranked list

The topics of the above described task are structured. This, when combined with our goal of structuring and organizing information to form 'answers', suggests the hypothesis: that *organizing information with regard to task structure is helpful to users*.

Intuitively, this makes sense. The goal of an interactive subject is to locate documents that pertain to as many different instances of the topic as possible. Given that there is no benefit¹ in locating documents that cover previously discovered topic instances, it would seem desirable to organize the candidate documents in such a way that documents addressing different instances of the query topic were separated into different groups. Ideally, the interactive user could then simply select a single representative document from each instance group. Further, these instance groups could help the user to organize the identified instances into a final answer.

How, therefore, should the candidate information be organized? The approach we chose to explore is clustering. The questions that we attempted to answer include: 1) Can users recognize good clusters? 2) Do users prefer a clustering approach? And 3) Are users more effective with a clustering approach?

We set up two experiments to answer these questions. In the first experiment, we recruited four subjects. Their task was to judge the relevance of a cluster to the topic based only on the description of cluster. As shown in Figure 1, the description of a cluster includes the ten highest-weighted terms from the cluster centroid, five most

¹ In fact, because Interactive Track experiments are conducted within a fixed time limit, it is counter-productive to locate or view documents that only address previously identified instances of a topic.

frequent word phrases from all documents in the cluster, and the title of the top three ranked documents in the cluster. This experiment shows that subjects were able to correctly determines from the cluster description which clusters were likely to contain relevant information, and which were not.

In the second experiment, we aimed to compare the clustering structure (as shown in Figure 3) with a ranked list on the question answering task [16]. This experiment was done in the TREC7 interactive track environment. We recruited sixteen subjects, each of them searched 4 topics on the cluster-based interface, and another 4 topics on the ranked list interface (similar to the design in Table 1). The result showed that overall there is no significant difference in terms of aspectual recall and aspectual precision between the cluster-based interface and the ranked list, although a fairly clear preference for the clustering structure was shown in subjects' comments as captured in questionnaires.

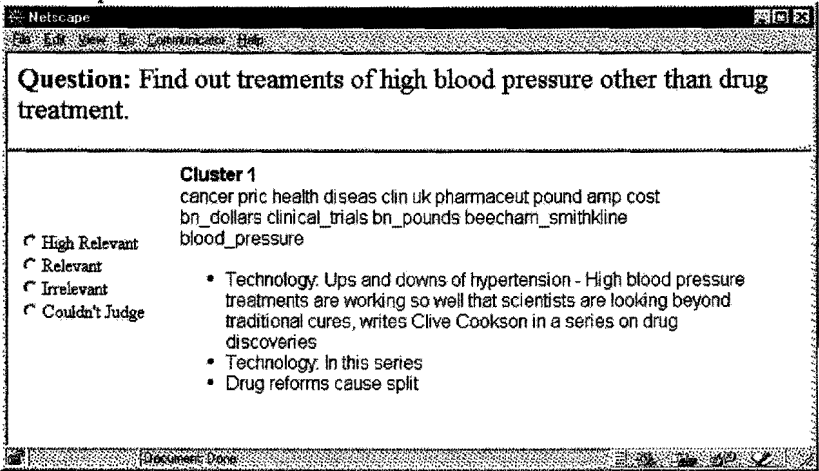


Figure 2. The interface for judging the relevance of clusters

Observations from Investigation 1:

- “Obvious” improvements often do not work. This has often been shown in information retrieval experiments, but extends to work involving users as well. This experiment is an example – there are others we have seen no performance gain.
- There may be a substantial difference between user performance and user preference. We have several times observed that users like a system that provides no performance gain. It is thus important to measure both, to understand the implications of a system variation.

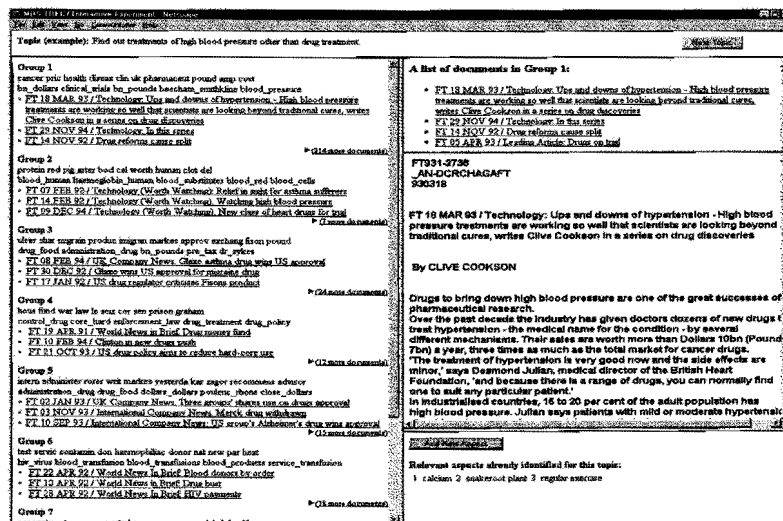


Figure 3. The cluster-based interface

Investigation 2: Task-biased summary versus generic document summary

We analyzed the logged data from the above second experiment, we did find that while the subjects could use the structured delivery format to locate the group of relevant documents, the subjects often either failed to identify a relevant document from the document surrogate or were unable to locate the answer component present in a relevant document. Thus we hypothesized that one of the causes for potential gains from the structured delivery not being realized is that in our test systems the tools used to differentiate the answer containing documents from non-answer containing documents were inadequate for the task of question answering. There is a lack of proper document preview mechanism for subjects to identify the answer-carrying documents.

In the previous experiment, the preview of a retrieved document is represented by its title. The title usually tells the main theme of the document, but very often an answer component (or an instance) exists only within a small chunk of the document, and this small chunk may not necessarily be related to the main theme of the document. For example, for the question "which was the last dynasty of China: Qing or Ming?", the titles of the first two documents are: "Claim Record Sale for Porcelain Ming Vase" and "Chinese Dish for Calligraphy Brushes Brings Record Price". The themes of the two documents are "Ming Vase" and "Chinese Dish" respectively, but there are two sentences in each document that mention the time range for Ming Dynasty and Qing Dynasty. By reading only the title, the subjects may miss a chance to find the answer easily and quickly, even the answer components are located in the top ranked documents.

So in our next experiment, we moved on to provide a user a task-biased document preview/summary. The research question we investigated in the experiment is: given a same list of retrieved documents, will the variation in document summaries improve user's performance on question answering task?

We conducted two experiments that compared two types of candidate lists in two experimental systems [17]. One system uses the document title and the first 20 words of a document as the document’s summary, while the other system uses the document title and the best three “answer indicative sentences” extracted from the documents as the document’s summary (as shown in Figure 4). The first experiment was done in the TREC9 interactive track environment (the documents in the collection are news articles). A second confirming experiment repeated the first experiment, but with different test collection (with web document collection) and subjects. The purpose of the second experiment was to confirm the strong results from the first experiment and to test whether the methodology could be generalized to web data.

Both experiments showed that subjects took less effort (issued fewer queries and read fewer documents), but found significantly more answers from the interface with task-biased document summary than the interface with the general document summary. Subjects also preferred the task-biased document summary. This result indicates that it makes difference by constructing a delivery interface that takes into account the nature of the task.

The set of experiments indicate that different search tasks may require different delivery methods. For the task of finding relevant documents, it has been found that under some circumstances the clustering of retrieved documents was better than a ranked list [1]. However, for the task of question answering, we didn’t find this delivery method performed better than a ranked list. While the second experiment indicates that a relatively simple document summary can significantly improve the searcher’s performance in question answering task.

Observations from Investigation 2:

- In both of these experiments, it appeared intuitively “obvious” that the test system (the clustering structure and the task-biased document summary) offered advantages over the base system (the ranked list). Only in the second case did the evidence back up the obvious.
- In this experiment we found a significant difference in performance. The raw average difference in performance was high – there was a 33% performance difference between the test system and the experimental system. This did give a statistically significant difference, but a smaller experiment, or an experiment where the difference was less extreme might have showed a difference that was not statistically significant. Using an ANOVA analysis below we can see that while the effect of the system is significant, so is the effect of the topic, and to a lesser extent, the user (our user population was chosen to be homogeneous).

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
System	1	1.205128	1.205128	16.54529	0.0001086
User	15	1.870922	0.124728	1.71240	0.0642876
Topic	7	5.234581	0.747797	10.26656	0.0000000
System:User	15	0.820926	0.054728	0.75137	0.7255804
System:Topic	7	0.518039	0.074006	1.01603	0.4262646
Residuals	82	5.972726	0.072838		

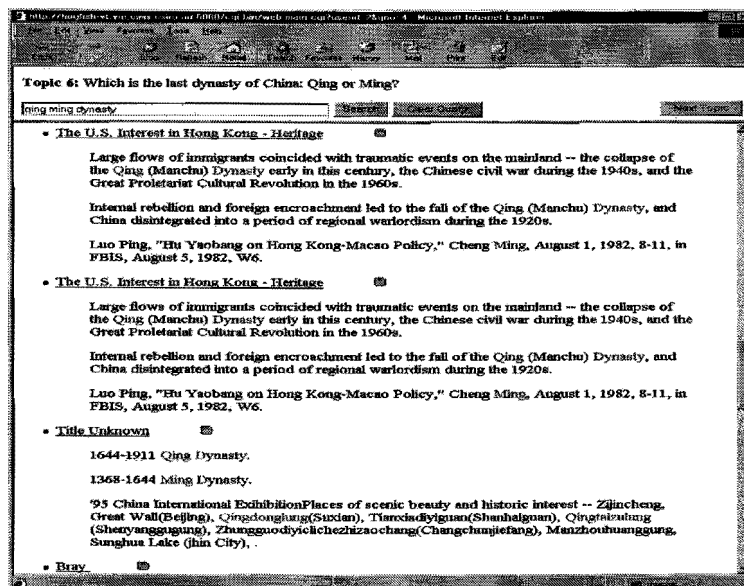


Figure 4. The interface for showing the answer indicative sentences.

3 Lesson Learned

From these experiments, and our observations of other research teams who have contributed to the TREC interactive track, we can make some comments. Firstly from a positive perspective:

- The TREC leverage the effort to build the evaluation platform – there is significant value to research teams working on the same task using the same data. What is learned by one group can be checked with other research teams' outcomes, so it is possible to identify potential trends, but also possible outliers.
- As we know the user involved evaluation is costly, each time only a couple of hypothesis can be tested. A common evaluation platform provides an opportunity for two or more research groups to co-ordinate their experiments in a way so that more hypotheses can be tested, as demonstrated in our current study with Rutger's team [18].
- A carefully constructed study using balanced experimental conditions such as the Latin square design can deal effectively with some of the possible interaction effects – observe the results for the interaction effects in the ANOVA analysis above showing little contribution.
- Effective delivery works in the right context – knowing the detail of the task can allow information to be organised and delivered more effectively than the standard list format of "normal" information retrieval systems

On the other hand:

- Even we had a common evaluation platform, it is hard to make direct comparisons of different systems across sites – there is often just too much variation, and in any case participating research groups were often exploring different hypotheses.
- It is hard to get a statistically significant difference, as the task variation and user variation often overwhelmed potential system variation leading to a lack of experimental power. It is well worth attempting to determine the variation of performance in a small experiment to determine whether there is any hope of achieving a result of statistical significance.
- Repeatability of experiments is difficult and expensive and may well not produce identical/consistent results, particularly given user variation.

It is important to measure both user performance on the task, and user preference. We have observed that these measures often do not correlate (in a short evaluation period and laboratory controlled setting). Yet both are important, if the target user population likes a system refinement, but it does not lead to productivity gains, it is questionable, but equally if performance improves, but the users do not like the approach, then there may well be a problem in uptake. (Maybe a longitudinal study is needed to observe the correlation between preference and the performance.)

4 Conclusions

We have described something of the history of evaluation in information retrieval from the perspective of the TREC interactive track. We then described some of the experiments that we have conducted, and discussed some of the lessons learned.

There are two key issues that we have often observed in interactive information retrieval. The first issue is that human preference is often not correlated with human performance. Consequently, evaluation that relies solely on either form of evaluation is not reliable. The second issue is that genuine improvements are very difficult to validate, as system variation tends to be dominated by task variation and user performance variation. Consequently, the statistical power of these experiments, often already very expensive to conduct because of user participation, can be quite low. Thus we argue for staged experiments where only very “obvious” system performance gains are explored.

In the end, simple performance measures have proved less helpful than deeper analysis of just how people interact with their information systems.

References

1. J. Allan, A. Leuski, R. Swan and D. Byrd. Evaluating combinations of ranked list and visualizations of inter-document similarity. *Information Processing and Management*, Vol. 37, Number 3, pp. 435-458, 2001.

2. N. J. Belkin, C. Cool, D. Kelly, S. J. Lin, S. Y. Park, J. Perez-Carballo and C. Sikora. Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing and Management*, Vol. 37, Number 3, pp. 403-434, 2001.
3. M. Hearst and J. O. Pedersen. Reexamining the cluster hypothesis: Scatter/Gather on retrieval results. In *Proceeding of the 19th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 76-84, 1996.
4. W., Hersh, A. Turpin, S. Price, D. Kraemer, B. Chan, L. Sacherek, D. Olson, Do batch and user evaluations give the same results? In *Proceedings of the 23rd Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.17-24, 2000.
5. P. Ingwersen. *Information Retrieval Interactions*. Taylor Graham, 1992.
6. K. Sparck-Jones. The cranfield tests. In K. Sparck-djones (editor), *Information Retrieval Experiment*, pp. 256-284. Butterworths, 1981
7. K. Sparck-Jones. Further relections on TREC. *Information Processing and management*, vol. 36, pp.37-85, 2000.
8. F. W. Lancaster. *Information Retrieval Systems: Characteristics, Testing and Evaluation*. New York: John Wiley & Sons, 1979.
9. R. Osdin, L. Ounis and R. W. White. Using hierarchical clustering and summarisation approaches for web retrieval: Glasgow at the TREC 2002 interactive track. In *Proceedings of TREC 2002*.
10. S. E. Robertson and M. Beaulieu. Research and evaluation in information retrieval. *Journal of Documentation*, Vol. 53, Number 1, pp.51-57, 1997.
11. G. Salton. The smart environment for retrieval system evaluation – advantages and problem areas. In Karen Sparck Jones (editor), *Information Retrieval Experiment*, pp. 316-329. Butterworths, 1981.
12. L. T. Su. The relevance of recall and precision in user evaluation. *JASIS*, vol.45, Number 3, pp.207-217, 1994.
13. J. Tague-Sutcliffe. Measuring the informativeness of a retrieval process. In *Proceddings of the 15th annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.23-36, 1992.
14. E. M. Voorhees and D. Harman. Overview of the fifth text retrieval conference (TREC-5). In *Proceedings of the fifth text retrieval conference*, pp.1-28, Gainthburg MD,USA, 1997.
15. P. Over. The TREC interactive track: an annotated bibliography. *Information Processing and Management*, Vol. 37, Number 3, pp.369-381, 2001.
16. M. Wu, M. Fuller and R. Wilkinson. Using clustering and classification approaches in interactive retrieval. *Information Processing and Management*, Vol. 37, Number 3, pp. 459-484, 2001.
17. M. Wu, M. Fuller and R. Wilkinson. Search performance in question answering. In *Proceedings of the 24st ACM SIGIR International Conference on Research and Development in Information Retrieval*. pp.375-381, 2001
18. M. Wu, G. Muresan, A. McLean, M. Tang, R. Wilkinson, Y. Li, H. Lee and N. Belkin. Human versus Machine in the Topic Distillation Task. To appear in *Proceedings of the 27st ACM SIGIR International Conference on Research and Development in Information Retrieval*. July 2004.