Intelligent Recommendation System for e-learning Platforms

Bruno Tavares
GECAD
ISEP
Rua Dr. António Bernardino de Almeida, 431, Porto
Portugal
bmtav@isep.ipp.pt

Luiz Faria
GECAD
ISEP
Rua Dr. António Bernardino de Almeida, 431, Porto
Portugal
lef@isep.ipp.pt

Constantino Martins
GECAD
ISEP
Rua Dr. António Bernardino de Almeida, 431, Porto
Portugal
acm@isep.ipp.pt

Abstract: As more and more digital resources are available, finding the appropriate document becomes harder. Thus, a new kind of tools, able to recommend the more appropriated resources according the user needs, becomes even more necessary. The current project implements an intelligent recommendation system for e-learning platforms. The recommendations are based on one hand, the performance of the user during the training process and on the other hand, the requests made by the user in the form of search queries. All information necessary for decision-making process of recommendation will be represented in the user model. This model will be updated throughout the target user interaction with the platform.

Key-Words: user modeling; user-adapted systems; recommending systems; e-learning systems; query clustering.

1 Introduction

The need to constant learning leads to an increased demand for e-learning systems. One of the greatest challenges in e-learning platforms is the ability to adapt to different users. In order to improve user experience in e-learning platforms, an Intelligent Recommendation System (IRS) is being developed as a module that can be used by a Learning Content Management System (LCMS). The development is been done specifically for this purpose and aims to increase the efficiency of the LCMS and provide a learning experience best suited for each individual.

This module is inserted as part of e-learning 3.0 project. The QREN 5635 – e-learning 3.0 is been developed in a collaboration effort by School of Engineering – Polytechnic of Porto (ISEP), University of Beira Interior (UBI), IZONE SGPS-SA and MAISIS – Information Systems.

The most important goal is to provide suggestions to the LCMS users. These suggestions are links to resources available on the LCMS and are compatible with the needs and preferences of the users. These needs and preferences will be defined through a user profile represented in a User Model (UM), which will be updated throughout the user interaction with the LCMS.

One of the requirements originally set for the LCMS was to develop mechanisms to provide intelligent support to users. The inclusion of these mechanisms targets the increase of training users efficiency, showing them the resources potentially more relevant [7][14]. The mechanism designed for this purpose is a module in the form of an Intelligent Recommendation System, able to recommend suitable resources to the user profile. These resources may correspond to SCORM packages, based on a repository, or other users of the platform, whose profiles and skills are appropriate to contribute to the formation process of the user concerned.

The rest of this paper is structured as follow: section 2 will present the problem formulation where will be approached problems that need to be addressed when developing an IRS. In section 3, the problem solution, it is suggested solutions to the problems previously identified. This section is divided on several sub-sections representing different parts of the solution. Section 4 explains how different recommendation techniques are integrated as a hybrid recommendation system. Finally conclusions are given in section 5.

2 Problem Formulation

Individualization is a key factor to achieve effective learning. This individualization or customization in e-learning systems refers to the ability to adapt to different learning rates and different knowledge background. In order to achieve this adaptation, the most advanced LCMS are using techniques born on the Intelligent Tutoring Systems (ITS) field. The application of these techniques tries to fill the role of tutor, aiming to recommend the most appropriate learning content to each user. In this project the intelligent tutoring system is designed as an IRS, for

ISBN: 978-1-61804-106-7

the IRS to appropriately must necessarily be able to evaluate the knowledge that each user has on different subjects and correctly suggest the next resource that the user must study [8][14]. Other's problems appearing when developing an IRS is the necessity to handle a lot of data, data with different types and forms, that needs to be considered and given different weights [3][16]. It is essential to submit the data to a pre-processing step, where it is necessary to address problems like words that are spelled the same way but have different meanings, in order to get better results. Dealing with a lot of resources where the information about those resources is limited or is not presented clearly can be a challenge and it is a problem that needs to be addressed. In the end it is necessary to have in mind that the resources change over time, new resources gets added, some get updated, upgraded, or other deprecated and removed.

In a world where education is shifting, where the need to continuous education is increasing every day, the need to access education on different levels and at different paces is reflected in the search for new techniques to help us getting the best strategy to satisfy our needs [19]. To be able to suggest the best resource for a specific user the problem was divided in two different types of learning, formal and non-formal.

3 Problem Solution

The IRS is being developed as a web service component based on Representational State Transfer (REST) protocol, like the remaining modules of the LCMS. The services available will allow other modules to perform updates on users profiles based on user interaction with the LCMS, and still get recommendations of resources. These recommendations may be obtained in two different contexts, formal learning context and non-formal learning context.

The generated recommendations will be based on the user's profile and search queries defined by the user. These types of information will be used to find resources whose meta-descriptions are compatible. To increase the chances of finding compatible resources, the terms used in the existing descriptions of the user profile, the queries, as well as the terms used in resource descriptions (metadata) will be processed using a technique of stemming.

Research in large repositories can be complex and demanding in terms of computing resources [23]. In order to simplify and streamline the process of searching for contents, the repository will be targeted by a clustering operation. The clustering operation will organize contents into groups or clusters, whose elements share similarities in terms of their descriptions. For this purpose it has been studied different clustering algorithms and their implementations. Were also studied using different metrics, like the Euclidean distance used to obtain measures of compatibility between objects [3][16].

Based on stemming and clustering operations, the recommendations will be generated through the implementation of techniques for collaborative filtering and content-based filtering [13][18]. These techniques allow the generation of personalized recommendations based on assessments made by other users with similar profiles on the recommended resources, as well as based on the preferences and needs. The resource evaluation carried out by users normally leads to a good indicator about the value of the resource to the users with similar profiles [13][23], so this information will be desirable for recommendation.

3.1 User model

There is a lot of information about the user that most platforms have but don't use to improve the user experience [4][5][6]. Be able to correctly collect user information is crucial. With a strong user model it is possible to create stereotypes with the goal to generate users groups. With that information it is possible to apply more weight into resources that users from the same group had already used successfully, leading to better recommendations [4][5][6]. In order to facilitate the user data structuring, the user model was subdivided into two areas, the Domain Independent Data (DID) and Domain Dependent Data (DDD) [4][5][6][17].

The DID aggregates the general information of the user that does not depend on any course, being this information used in a context of informal learning [17].

The DDD aims to add all the user information that is somehow associated with a course in order to focus the recommendations in a context of formal learning [17].

3.2 Learning contexts in e-learning 3.0

There are two different learning contexts, the formal learning context and the non-formal learning context. The learning context indicates the kind of environment in which the request for recommendation is inserted, influencing the behavior of the IRS.

In the formal learning context, the user is registered in a course, which has an associated curriculum (or script) previously defined. The recommendations take into account the progression of the user and the user model, where the suggestions are significantly influenced by the DDD. In the formal learning context the recommendations need to help the user to complete the course requirements.

In the context of non-formal learning process, the user has the initiative to consult existing resources, without a training plan pre-defined. In formal learning context there is a more strict orientation, while in the context of non-formal that rigidness disappears allowing the IRS to recommend different types of resources that might be relevant to the user. With this kind of approach it is possible to use other type of information that otherwise will probably never be taken into consideration. That information can be found in the user model DID under classes such as social data and professional data, and can be used by the recommendation system not only to create stereotypes but also to indicate user's preferences.

3.3 Stemming

The stemming process consists in extracting words to their base element or radical, which defines the meaning of the word. The application of this process in an early step can increase the success of the research. For this purpose stemming algorithms and their implementations dedicated to different languages were analyzed [24][26]. This process is used to clear the inserted data removing extra-noise, such as dropping elements as plurals and gerunds, and increasing the similarity between words.

3.4 SCORM

The management of large repositories can be a complex task. Standardization of resources information will allow us to simplify this task. Been Shareable Content Object Reference Model (SCORM) a extensively used norm collection, adopting this standard not only help us to manage the system resources but also bring us one step closer to other's systems that already use that norm and want to integrate our IRS [25].

In addition, the SCORM compliant packages include a meta-description file, using the Learning Object Metadata (LOM) standard, which allow us to access the description of the package contents [9].

The resources recommended by the IRS will be under SCORM standard.

3.5 Recommendation techniques

The use of hybrid systems allows a greater flexibility to refine the behavior of our system, making it possible to add or remove recommendation techniques, increasing the IRS flexibility and adaptability, therefore increasing the possibility of the system evolution.

The hybrid IRS is composed of a clustering algorithm, a similarity algorithm, a content-filtering technique, a collaborative filtering technique and a resource classification tool.

3.5.1 Clustering

The search engines' importance has increased to levels where users demand for higher accuracy on query's results. This demand has stimulated the development of information retrieval and search clustering techniques. Initially were analyzed forms of clustering to allow unanticipated results. This analysis led to consider self-organizing maps, Kohonen networks and DBSCAN [2][10][11][12][15].

DBSCAN is a density-based clustering algorithm with the particularity of not needing to define the number of initial clusters. Like others clustering algorithms, in DBSCAN is not possible to evaluate the quality of the clusters results [10]. DBSCAN strongly depends on the quality of the distance measure or similarity.

3.5.2 Similarity

The similarity is the distance between the clusters objects. Been able to correctly calculate similarity between each object will positively influence the results of the cluster algorithm, and the resources recommended [3][16].

3.5.3 Content-filtering technique

Content-filtering techniques are extensively used for filtering data. In this technique the resources are blocked or allowed based on the analysis of the resource meta-descriptions. This technique reorders the elements of the cluster giving greater importance to the most popular content. This way, it is intended to give greater rank to the content that is more widely used.

3.5.4 Collaborative filtering technique

Collaborative filtering is a technique used to filter information based on feedback provided by users. For example, if a user likes a new resource, users

ISBN: 978-1-61804-106-7

that in past liked the same resources will more likely like the new resource. Using this technique it is intended to incorporate feedback given by the user via the resource classification tool in the process of recommendation.

3.5.5 Resource classification tool

To improve the collaborative filtering algorithm it is necessary to use some sort of classification mechanism where the users can provide some feedback about the resource they are using at the moment [13]. For resource classification the IRS is prepared to consider inputs, such as like or dislike tags about a resource, in the process of generating recommendations.

4 Hybrid recommendation system

When a user asks for a recommendation the query's keywords are matched with a cluster and then the cluster is re-organized using collaborative filtering and content-based filtering techniques [18], the reorganization of the cluster aims to sort the elements of the cluster, positioning the contents with greater relevance in the first place. The content-based filtering techniques analyze the resources meta-descriptions to identify resources best suited for the user. The collaborative filtering consists in the analysis of users behaviors, using the information gathered it is possible to reorganize the cluster resources, indicating which resources is the more popular.

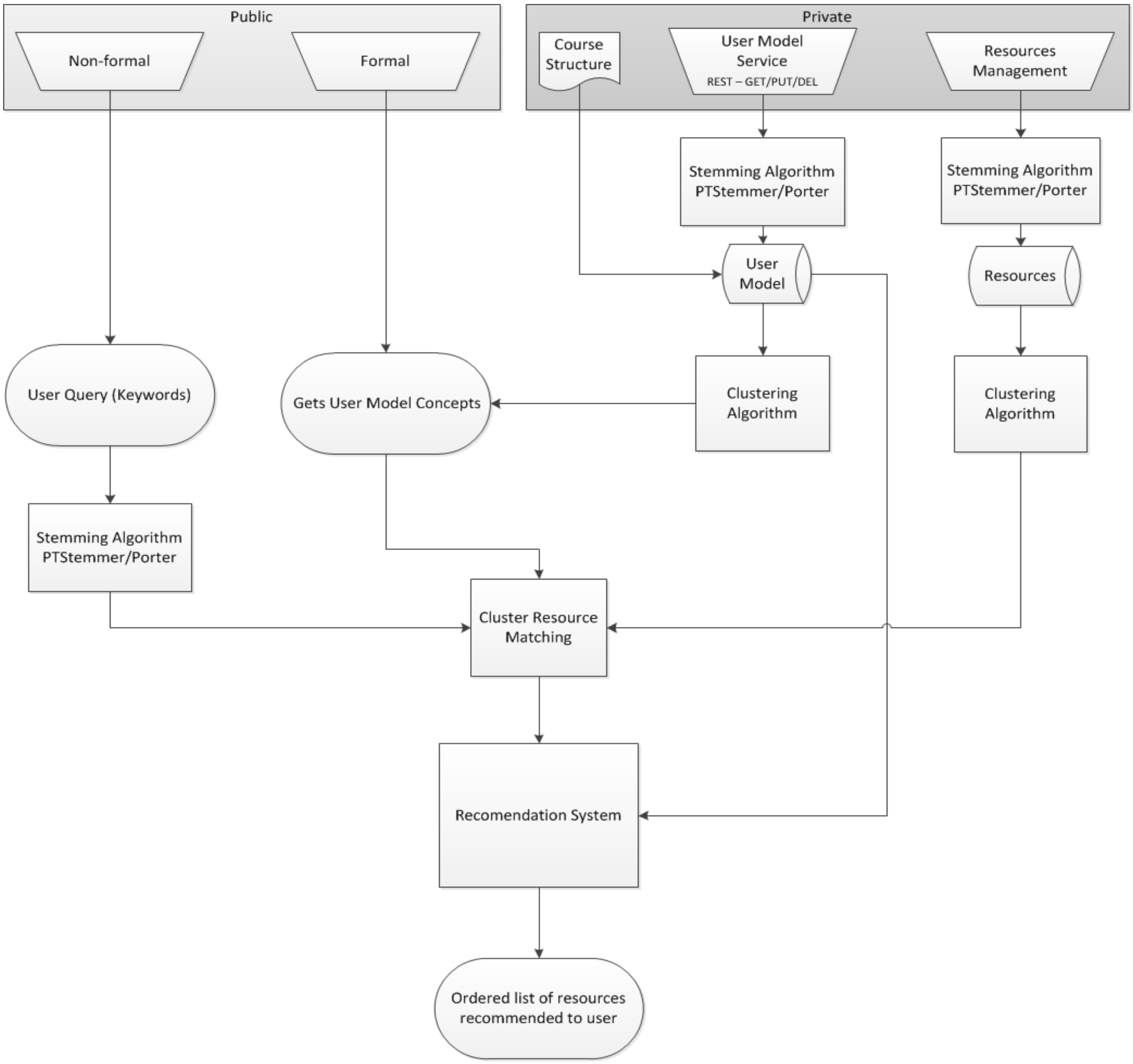

Fig. 1 – Recommendation scheme

Fig.1 illustrate part of the IRS that is been developed, where is possible to see the desired flow of recommendations and how some parts of the recommendation algorithm are interconnected.

The IRS has been divided into two parts, public and private. The public part publishes interfaces to provide public services used to forward responses to the LCMS users and also gather users feedback. The private part of IRS provides, to the LCMS, the services necessary for the management of the system, as well as services to manage the necessary data. These services should only be connected with the administrative part of the LCMS.

The resources management can be made with the use of an external service that provides data within existing resources. The resources update can be made using a procedure such as "cron job". Or alternatively, the LCMS can manage existing information about resources in the database of the IRS delegating this responsibility to the LCMS. The LCMS can use the resources services available to manage the learning resources information.

In the resource management service, it should be understood that it will only be used to the management of the information about the resources, the resources meta-description and an indication of where the resources are stored. In the user model is registered what are the courses where the user is or has been enrolled. The structure of the course, or curriculum, is represented in the course structure.

The clustering algorithm is used not only to group resources but it is also used to group users. The clustering of users is based on the information contained in the user model, such as the user's competences or skills, and aims to group similar users. Thus, in the step where the user model concepts are obtained, "*Gets User Model Concepts*", it is possible to expand the user model concepts. The user model concepts expansion aims to increase the information pool of one user, using information from similar users. With a larger information pool it is possible to extend the recommendations range.

All input data are treated using the stemming algorithm. This procedure will reduce the words to its radical form. Similar words or words from the same family are reduced to the same word, reducing the noise existing in the data.

In the matching step "*cluster resource matching*" is carried out the matching between the keywords entered by the user in the context of non-formal learning, and the cluster resource that presents greater proximity. The same matching step is used to perform the matching between the concepts obtained in the user model, in a formal learning context, and the cluster resource that presents greater proximity.

In the recommendation system is carried out filtering and sorting of the elements of the cluster, using the collaborative filtering and content-based filtering techniques. In this last step it is possible to introduce new filtering algorithms that may emerge.

The system also considers if the referenced resources have been previously viewed by the user. Suggesting only contents that have not been seen by the user gives greater relevance to other contents, thus, study alternatives are always suggested.

It is being considered the possibility to give more weight to the most popular keywords. The objective of this measure is to influence the IRS to give greater importance to issues related to the educational environment in which the LCMS is inserted. Combining these technique's it's possible to suggest an ordered list of recommendations to a user where the first resource in the list is the most recommended one.

5 Conclusion

The recommendation system devised, which scheme is presented in Fig. 1, is composed by a sequence of modules implementing several techniques. The application of these techniques in an isolated manner to other kind of problems was already been proved. However, our goal is to prove that these techniques can be used together to increase the behavior of the recommendation systems. By combining several solutions already tested it is possible to come up with a solution to our problem.

The use of ontologies to expand the keywords, extending the keywords pool, was considered. The initial idea was to be able to get more results for smaller queries [1][20][21]. Using ontologies for queries expansion wasn't a viable solution, because of the need to create and handled a universal ontology that is able to adapt to different e-learning environments. On another hand, from analyzing tagging recommendation, widely used on social sharing web-sites, techniques like collaborative filtering and content-based filtering to recommend resources to users were identified [13][25].

Using a scalable hybrid algorithm opens the possibility to upgrade and update our algorithm with new techniques than can emerge. For example, it is being analyze the possibility to introduce in the future the option to boost new resources, so that new resources can move to the top quicker in order to be viewed and classified by users.

6 Acknowledgments

The authors would like thank FEDER, POFC and QREN for their support to GECAD unit and the 5635 – e-learning 3.0 project.

References:

[1] Andreou A, *Ontologies and Query expansion*, School of Informatics University of Edinburgh, 2005.

[2] Cail D, He X, Li Z, Ma W, Wen J, *Hierarchical Clustering of www Image Search Results Using Visual, Textual and Link Information, 2004*.

[3] Cattuto C, Benz D, Hotho A, Stumme G, *Semantic Analysis of Tag Similarity Measures in Collaborative Tagging Systems, 2008*.

[4] Coelho B, Figueiredo A, Martins C, *Tours Planning Decision Support,* GECAD – Knowledge Engineering and Decision Support Group, Computer Science Department, Institute of Engineering – Polytechnic of Porto, Porto, Portugal, 2009.

[5] Coelho B, Martins C, Almeida A, *Adaptive Tourism Modeling and Socialization System*, GECAD – Knowledge Engineering and Decision Support Group, Computer Science Department, Institute of Engineering – Polytechnic of Porto, Porto, Portugal, 2009.

[6] Coelho B, Martins C, Figueiredo A, *Hybrid Architecture for Tourist Modeling and Recommender System*, GECAD – Knowledge Engineering and Decision Support Group, Computer Science Department, Institute of Engineering – Polytechnic of Porto, Porto, Portugal.

[7] Freedman R, *What is an Intelligent Tutoring System?*, Northern Illinois University. Fall 2000.

[8] Gamboa H, Fred A, *Designing Intelligent Tutoring Systems: a bayesian approach, 2003*.

[9] IEEE, Learning Technology Standards Committee, *WG12: Learning Object Metadata*, http://ltsc.ieee.org/wg12/, December 2002.

[10] Ji-Rong Wen, Hong-Jiang Zhang, *Query Clustering in the Web Context*, Microsoft Research Asia 3F, Beijing Sigma Center No. 49, Zhichun Road, Haidian District Beijing, P.R. China 100080, December 2003.

[11] Ji-Rong Wen, Jian-Yun Nie, Hong-Jiang Zhang, *Query Clustering Using User Logs, January 2002*.

[12] Ji-Rong Wen, Jian-Yun Nie, Hong-Jiang Zhang, *Clustering User Queries of a Search Engine, May 2001*.

[13] J¨aschke R, Marinho L, Hotho A, Schmidt-Thieme L, Stumm G, *Tag Recommendations in Folksonomies, 2007*.

[14] Lane H C, *Intelligent Tutoring Systems: Prospects for Guided Practice and Efficient Learning*, Institute for Creative Technologies University of Southern California. July 27, 2006.

[15] Limam L, Coquil D, Kosch H, Brunie L, *Extracting user interests from search query logs: A clustering approach, 2010*.

[16] Markines B, Cattuto C, Menczer F, Benz D, Hotho A, Stumme G, *Evaluating Similarity Measures for Emergent Semantics of Social Tagging, 2009*.

[17] Martins C, Faria L, Carvalho C V, Carrapatoso E, *User modeling in adaptive hypermedia educational systems, Educational Technology & Society*, 11(1), 2008, pp. 194-207.

[18] Pazzani M J, Billsus D, *Content-Based Recommendation Systems, 2007*.

[19] Rane A K, *Intelligent Tutoring System For Marathi*, Karnataka State Open University Mysore, December 2005.

[20] Sally M. El-Ghamrawy, Ali I. EI-Desouky, M. Sherief, *Dynamic Ontology Mapping for Communication in Distributed Multi-Agent Intelligent System, 2009*.

[21] Siddharth Kaza, Hsinchun Chen, *Evaluating ontology mapping techniques: An experiment in public safety information sharing*, Department of Management Information Systems, University of Arizona, United States, December 15, 2007.

[22] Smith K A, Alan Ng, *Web page clustering using a self-organizing map of user navigation patterns*, School of Business Systems, Monash University, P.O. Box 63B, Victoria 3800, Australia, 2003.

[23] Song Y, Zhang L, C. Lee Giles, *Automatic Tag Recommendation Algorithms for Social Recommender Systems*, The Pennsylvania State University, 2008.

[24] Smirnov I, *Overview of Stemming Algorithms*, DePaul University, December 2008.

[25] Tzone I Wang, Kun Hua Tsai, Ming Che Lee and Ti Kai Chiu, *Personalized Learning Objects Recommendation based on the Semantic-Aware Discovery and the Learner Preference Pattern*, National Chung Kung University, Taiwan, 2007.

[26] Willett P, *The Porter stemming algorithm: then and now*. Program: electronic library and information systems, 40 (3). pp. 219-223, 2006.