Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing

Suzanne Little[1], Joost Geurts[2], and Jane Hunter[3]

[1] ITEE Dept. University of Queensland little@itee.uq.edu.au
[2] CWI, Amsterdam joost.geurts@cwi.nl
[3] DSTC, Brisbane jane@dstc.edu.au

Abstract. This paper first proposes a high-level architecture for semi-automatically generating multimedia presentations by combining semantic inferencing with multimedia presentation generation tools. It then describes a system, based on this architecture, which was developed as a service to run over OAI archives - but is applicable to any repositories containing mixed-media resources described using Dublin Core. By applying an iterative sequence of searches across the Dublin Core metadata, published by the OAI data providers, semantic relationships can be inferred between the mixed-media objects which are retrieved. Using predefined mapping rules, these semantic relationships are then mapped to spatial and temporal relationships between the objects. The spatial and temporal relationships are expressed within SMIL files which can be replayed as multimedia presentations. Our underlying hypothesis is that by using automated computer processing of metadata to organize and combine semantically-related objects within multimedia presentations, the system may be able to generate new knowledge by exposing previously unrecognized connections. In addition, the use of multilayered information-rich multimedia to present the results, enables faster and easier information browsing, analysis, interpretation and deduction by the end-user.

1 Introduction and Objectives

Information about a particular person or topic can be created by multiple users, served by various services and dispersed across multiple sites over the Internet. Adoption of standardized metadata vocabularies and ontologies, expressed in standardized machine-processable languages such as the Resource Description Framework [26] or DAML+OIL [23] are contributing to the realization of the next generation Web - the Semantic Web. One of the key promises of the Semantic Web [1] is that it will provide the necessary infrastructure for enabling services and applications on the Web to automatically aggregate and integrate information into a sum which is greater than the individual parts.

The current Web technology is at the 'hunter-gatherer' stage. The result of a typical search is a sequential list of URL's, referring to the HTML pages which match the metadata search field, displayed according to rank. The fact that there are semantic relations between the retrieved objects or that many more semantically related information objects exist, is ignored in the final presentation of results.

In parallel with advancements in the development of the Semantic Web, is a rapid increase in the size and range of multimedia resources being added to the Web. Archives, museums and libraries are making enormous contributions to the amount of multimedia on the Internet through the digitization and online publication of their photographic, audio, film and video collections.

Within this paper we attempt to exploit all of these developments - the rapid growth in multimedia content, the standardization of content description and the semantic web infrastructure - to develop a system which will automatically retrieve and aggregate semantically related multimedia objects and generate intelligent multimedia presentations on a particular topic.

Using the Open Archives Initiative (OAI) [4] as a testbed, we have developed a service which uses the Dublin Core metadata published by the OAI data providers, to infer semantic relations between mixed-media objects distributed across the archives. Using predefined mapping rules, these semantic relationships are then mapped to spatial and temporal relationships between the objects. The spatial and temporal relationships are expressed within SMIL files which can be replayed as multimedia presentations.

Our premise is that by using automated computer processing of metadata to organize and combine semantically-related objects within multimedia presentations, the system may be able to generate new knowledge, not explicitly recorded, by inferring and exposing previously unrecognized connections. In addition, the use of multilayered information-rich multimedia to present the results, enables faster and easier information browsing, analysis, interpretation and deduction by the end-user.

The remainder of the paper is structured as follows. The next section describes related initiatives and projects and outlines how the work described here differs or builds on these. Section 3 describes the high-level system architecture, followed by details of the components and processes which make up our actual implementation. Section 4 provides the results of running a real example query and Section 5 concludes with a discussion of problem issues and future work.

2 Background

2.1 OAI

The Open Archives Initiative (OAI) [4] is a community that has defined an interoperability framework, the Open Archives Metadata Harvesting Protocol [3], to facilitate the sharing of metadata. Using this protocol, *data providers* are able to make metadata about their collections available for harvesting through an HTTP-based protocol. *Service providers* then use this metadata to create value added services.

To facilitate interoperability, data providers are required to supply metadata which complies to a common schema, the unqualified Dublin Core Metadata Element Set [2]. Additional schemas are also allowed and are distinguished through the use of a metadata prefix.

To date, OAI service providers have mostly developed simple search and retrieval services [5]. These include *Arc*, *citebaseSearch* and *my.OAI*. One of the more interesting services is *DP9*, a gateway service which allows traditional web search engines (e.g., Google) to index otherwise hidden information from OAI archives. Although originating in the E-Print community, OAI data providers now include multimedia collections such as the Library of Congress: American Memory collection [15], OpenVideo [17] and University of Illinois historical images [16]. Our goal is to exploit this increasing availability of multimedia resources and associated metadata to develop more interesting search, retrieval and aggregation services.

2.2 The Semantic Web

The Semantic Web [25] is an activity of the W3C which aims to extend the current Web by providing tools that enable resources on the web to be defined and semantically linked in a way that facilitates automated discovery, aggregation and re-use across various applications.

One of the cornerstones of the Semantic Web is the Resource Description Framework (RDF) [26]. RDF provides a common underlying framework for supporting semantically-rich descriptions, or metadata which enables interoperable XML data exchange. The Web Ontology Working Group are currently developing a Web Ontology Language (OWL), based on RDF Schema [28] (and its extension, DAML+OIL [23]) for defining structured, Web-based ontologies which will provide richer integration and interoperability of data among descriptive communities.

As the Semantic Web expands and more communities use RDF Schema or DAML+OIL to describe and annotate their data and build ontologies, the need for RDF-enabled storage and querying mechanisms to support large-scale Semantic Web applications, has grown. Research groups have been developing new query languages such as RQL [7] and SquishQL [14] which enable the storing and querying of web-based ontology/metadata standards, such as Dublin Core, RDF, RDFS, Topic Maps, DAML+OIL and the forthcoming Web Ontology Language (OWL).

A few of these RDF-enabled query languages (those based on logic models) also provide support for inferencing - in the form of either arbitrary deduction rules or user-defined inference rules. These rules enable new associations and knowledge, not explicitly recorded, to be inferred. Both Intellidimension's RDFQL [11] and TRIPLE [20] provide support for user-defined inference rules of the form *if A is the case then so is B*, to infer new knowledge and enable powerful deductive searches. A number of RDF and Topic Map inference engines (such as Cerebra and Empolis K2) exist, which also allow the automatic inferencing of new associations based on pre-defined rules.

A number of research groups are working on the next stage beyond semantic inferencing - the aggregation of semantically related data sources. Intellidimension's *RDF Gateway* [11] uses the new associations deduced from the RDFQL inferencing capabilities, to compile multiple data sources into a single knowledge base. *WebScripter* [10] is a tool that enables users to assemble reports by extracting and fusing information from multiple, heterogeneous DAML-ized Web sources. A number of research groups have used RST (Rhetorical Structure Theory) relations to link mixed-media resources into tree structures which can be translated into a coherent multimedia presentations [6], [24]. However, the idea of semantically inferring relationships between mixed-media resources on the fly, and then translating these relationships to generate multimedia presentations, is a relatively new idea and, will provide better user interfaces to integrated information sources.

2.3 SMIL, Cuypers

Synchronized Multimedia Integration Language (SMIL) [27] is a W3C Recommendation designed for choreographing web-based multimedia presentations which combine audio, video, text and graphics in real-time. It uses a simple XML-based markup language, similar to HTML, which enables an author to describe the temporal behavior of a multimedia presentation, associate hyperlinks with media objects and describe the layout of the presentation on a screen. Its advantages include platform independence, network and client adaptability and the simplicity of XML for generation.

Cuypers [24] is a research prototype system, developed to experiment with the generation of Web-based presentations as an interface to semi-structured multimedia databases. Given a rhetorical structure (which models the intended message of the presentation) and a set of rules (which map rhetorical structures to spatio-temporal relations). Cuypers generates a presentation that adheres to the limitations of the target platform (system capabilities) and supports the user's preferences. Further details of Cuypers are provided in Section 3.5.

2.4 Other Related Work

Some earlier research has focussed on using hyperlinks to link semantically related information objects [22] within hypertext documents. A certain amount of work has also been done on formalising semantic relationships between media items. Both the MAVIS-2 [21] project, and MPEG-7 Semantic Description tools [12] propose a separation of the semantic layer from the actual media content. The MAVIS-2 project expresses the semantics of the multimedia content in an ontology which has links from concept definitions to the media that represent them. Semantic relations between seperate media objects can be inferred if they are all described using the same ontology. Similarly, the MPEG-7 Semantic Description Tools define XML Schemas for describing the semantics of multimedia through a number of top level semantic entities such as Objects, Agents, Events, Concepts, State, Place, Time and SemanticRelation. As with MAVIS-2, the semantic descriptions are linked to the corresponding media segments through temporal and spatial media locators.

In addition to the need to formalize the semantic descriptions of multimedia, is the need to formally describe the segmentation metadata and low level audio and visual features/descriptors and formatting metadata. MPEG-7 has been designed specifically to standardize such multimedia content descriptors. In the future, as larger collections of multimedia described using MPEG-7 are developed, then we anticipate developing systems which can infer much richer semantic relationships through the MPEG-7 metadata and the associated MPEG-7 Ontology [9], but for the moment and in the context of this paper, we are limited to semantic inferencing using simple Dublin Core.

3 Implementation

3.1 Architecture

Figure 1 shows the high-level system architecture and the processes which transform the data models at each stage of the system. There are five stages in the overall presentation generation process. These are described in more detail in the following subsections:

- Iterative Search Process - users are able to interactively search and navigate the content of selected OAI archives, interpreting the retrieved results, selecting the pertinent resources and directing the subsequent search focus;
- Semantic Inferencing - the Dublin Core metadata is used to infer semantic relationships between the retrieved media objects;
- Mapping - the inferred semantic relationships are then mapped to spatial and temporal relationships or multimedia formatting objects[4](MFO)
- Presentation Generation - a multimedia presentation is generated from the input media objects, semantic relations, mappings to MFO's and other constraints;
- User-directed Presentation Regenerator - a user can change the focus of the current presentation and generate a new presentation by clicking on a media item of interest in the current presentation.

[4] MFO's are called communicative devices in Cuypers.

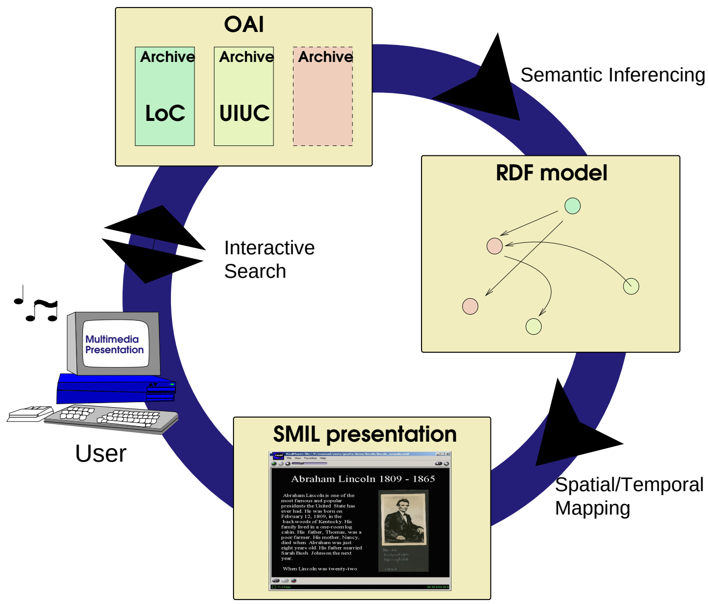


Fig. 1. Architecture overview

3.2 Interactive Search Process

Performing a simple keyword search on the metadata from the OAI archives is likely to return a large set of items, including much irrelevant information. For example, a search for *Lincoln* will return information about both the American President and the town in Nebraska. The result set is also limited, in that there are often many other interesting related resources which are missed by a traditional keyword search e.g., information about Stephen A. Douglas who argued against Lincoln in the slavery debate, would not be retrieved.

To overcome these restrictions and fully exploit the resources in the OAI, an iterative search process has been developed. This process manipulates the Dublin Core elements and values at each search stage to enable the discovery of semantically linked media items. The user first enters a keyword which describes their topic of interest. The system then performs a search across the *dc.title* and *dc.subject* elements of the archives' metadata to find occurrences of that keyword and presents the results to the user in a traditional list format. The user then has the option to direct the search by selecting those media items which are most pertinent to their area of interest. Based on these choices the system applies the semantic inference rules described in Section 3.3 to generate new searches based on the metadata of the chosen resources. For example, the system may take a *dc.contributor* value for a chosen resource and search for resources which have *dc.subject* equivalent to this value, in order to find out more about the colleagues of the creator of the original resource. Through such iterative processing, semantic relationships between resources related to the user's topic can be inferred. The user has the opportunity to direct the search after each retrieval of media therefore ensuring that the presentation maintains its focus and relevance to the user. An outline of the overall process is shown in figure 2.

Consider, for example, the two media items and their corresponding metadata, displayed in Figure 3. A simple keyword search for *Lincoln* will return item 1 but not item 2, a portrait

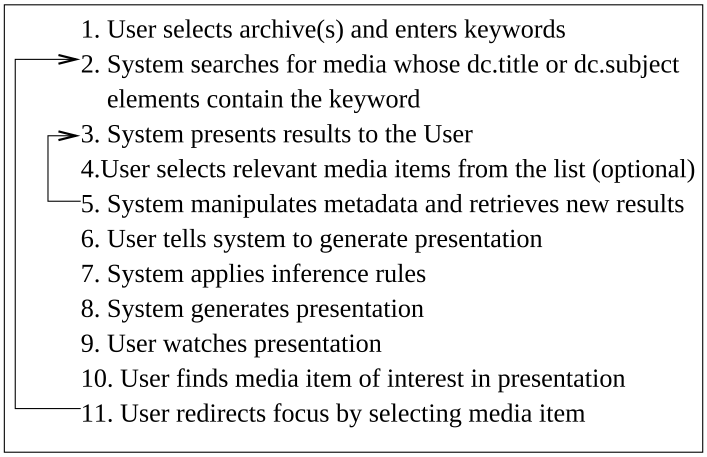


Fig. 2. Iterative Search Process

of Stephen Douglas, because the key term *Lincoln* doesn't appear in the metadata of item 2. Douglas was a contemporary and influential figure of Lincoln's and therefore item 2 is highly relevant to the topic *Lincoln*.

Using the iterative search process described here, the system takes the *dc.contributor* value associated with the first item, the advertisement for the 'Lincoln-Douglas' debate, and searches for occurrences of 'Douglas, Stephen' in the *dc.title* or *dc.subject* fields of resources. This retrieves the portrait of Douglas and, if selected by the user, it will be included in the final presentation.

Through such manipulation of the contents of the DC elements, it is possible to retrieve interesting, related media items that would otherwise be lost and hence build up a network of rich semantic relationships through a semantic inferencing process which is described in the next section.

3.3 Semantic Inferencing using Dublin Core

A set of pre-defined rules is applied to the Dublin Core metadata associated with the acquired set of media objects, to determine the semantic relationships between them. An example of such a relation is the *created* relation which can be inferred between two resources if, for example the *dc.creator* value of one resource equals the *dc.subject* value of a second image resource[5].

[5] This rule does not work in all cases and one of our goals is to determine the specific circumstances in which semantic inferencing works


```
<title>Lincoln-Douglas debate</title>

<subject>Abraham Lincoln</subject>
<subject>Stephen Douglas</subject>
<description>poster advertising re-enactment
of the Lincoln-Douglas Debate</description>

<date>Aug 27 1858</date>
<contributer>Abraham Lincoln</contributor>
<contributer>Stephen Douglas</contributor>
<type>image</type>
```

```
<title>Stephen Arnold Douglas,
portrait</title>
<subject>Stephen Douglas</subject>

<description>Stephen Arnold Douglas,
head-and-shoulders portrait, slightly
to left, facing light</description>
<date>1844-1860</date>


<type>image</type>
```

Fig. 3. Media items with selected metadata.

In order to define consistent mappings from semantic relations to spatial/temporal relations, we require a fixed set of semantic relations. In the context of our application, the MPEG-7 Semantic Relations Description Scheme specified in the MPEG-7 Multimedia Description Schemes specification [12] provides a good base set. Rather than restrict the allowed semantic relations to this set, we have chosen instead to define an object-oriented semantic relation ontology, which has the MPEG-7 semantic relations at the top level. This allows low-level domain-specific semantic relations to be inferred and assuming that they are a subproperty of a higher level MPEG-7 relation/property, which has a corresponding presentation construct, then we will be able to determine a spatial-temporal mapping (e.g., *translationOf* is a subPropertyOf *mpeg7:versionOf* and hence has the same spatial-temporal mapping).

Table 1 shows a list of inferencing rules which we have applied to the metadata of retrieved, selected resources - both to determine new searches and to infer specific semantic relations between resources retrieved at each stage. To date our work has mainly focussed on searches on people, the objects they have created and the influences on them, e.g., Lincoln, Picasso, Ellington. This has been both because of the nature of the multimedia content in the OAI archives and because of the nature of Dublin Core metadata.

The definition of equality in the inference rules(table reftbl:infrules is rather loose since the DC element values may not be exactly equal but may include sub-strings or perhaps eventually synonyms. By comparing combinations of element values more interesting semantic relationships (eg: *shareContext*, *versionOf*) can be inferred. Sequences or groups of media

Table 1. Example Inferencing Rules

1. IF (obj1[dc.creator] = obj2[dc.subject] AND (obj1[dc.creator] = obj2[dc.title] AND obj2[dc.type] = 'image') → created(obj2,obj1)
2. IF (obj1[dc.title] = obj2[dc.subject] AND obj2[dc.type] = 'text') → describes(obj2, obj1)
3. IF (obj1[dc.title] = obj2[dc.subject] AND obj2[dc.type] = 'image') → depicts(obj2, obj1)
4. IF (obj1[dc.source] = obj2[dc.identifier]) → sourceOf(obj2, obj1)
5. IF (obj1[dc.source] = obj2[dc.source] → shareSource(obj1, obj2)
6. IF (obj1[dc.title,dc.creator,dc.subject] = obj2[dc.title,dc.creator,dc.subject]) → versionOf(obj1, obj2)
7. IF ((obj1[dc.subject] = obj2[dc.subject]) AND (obj1[dc.date] = obj2[dc.date] OR obj1[dc.coverage] = obj2[dc.coverage] OR obj1[dc.contributor] = obj2[dc.contributor])) → shareContext(obj1, obj2)
8. IF (obj1[dc.creator] = obj2[dc.subject] AND obj2[dc.type] = 'image' AND obj1[dc.contributor] = obj3[dc.subject] AND obj3[dc.type] = 'image') → colleagueOf(obj2,obj3) AND colleagueOf(obj3,obj2)[6]
9. IF (versionOf(obj1, obj2) AND obj1[dc.date] < obj2[dc.date]) → precedes(obj1, obj2) AND follows(obj2, obj1)
10. IF (created(obj1,obj2) AND created(obj1,obj3) AND obj2[dc.date] < obj3[dc.date]) → precedes(obj2,obj3) AND follows(obj3,obj2)

items, related to a common item by the same relation, are ordered through *precedes* and *follows* relations, two examples of which have been provided here.

Rule 1 states that if an object's creator is the subject of another object which is also an image then the second subject is the 'creator' of the first. Figure 5 in Section 4 provides an example of the application of this rule. The photograph on the left shows Lincoln who is the creator of the speech which is represented by the image on the right. Because the *created* relationship is inferred for a number of portraits of Lincoln, a *group* relation is then inferred over the portraits.

The rules provided here are a relatively simple set of inferences based on binary relations and which focus primarily on 'people' searches. This has, to a certain extent, been because we are limited to simple Dublin Core metadata. Richer semantic relationships could be inferred through n-ary relations or more complex metadata, such as MPEG-7. This higher level of complexity would best be addressed using more powerful inference engines, such as RDFGateway [11], TRIPLE [20] or Cerebra. However, for simple, unqualified Dublin Core the rules listed here are adequate and can produce a structured, ordered results set suitable for translating into a presentation.

3.4 Mapping from Semantic Relationships to Temporal and Spatial Relationships

As the number of possible semantic relations is infinite, while the number of possible spatial and temporal relations are limited, we use a semantic relationship ontology/hierarchy in which all semantic relationships are derived from the top-level MPEG-7 semantic relationships. Table 2 shows the mapping from our inferred semantic relationships to the top-level MPEG-7 semantic relationships. If more complex domain-specific semantic relationships can be inferred then their mapping to MPEG-7 semantic relationships will need to be added to this table. Table 3 illustrates the corresponding mapping from the MPEG-7 semantic relationships to

Cuypers MFO's, desribed here as logical spatial and temporal relationships. For example, the fact that there is an order between created works can be illustrated by presenting them in chronological order. Similarly, a way of conveying 'grouping' between media items is to align them together spatially.

Table 2. Semantic Relations Mapping Table

Semantic	MPEG7/Parent
X created Y	X result Y
X colleagueOf Y	X accompanier Y
X depicts Y	X depicts Y
X describes Y	X annotates Y
X shareContext Y	X similar Y
X sourceOf Y	X component Y
X versionOf Y	X identifier Y
X precedes Y	X before Y
X follows Y	X after Y

Table 3. Mapping from MPEG-7 Semantic Relations to Spatio-temporal Relations

MPEG7/Parent	Examples of Temporal/Spatial Relations
X result Y	spatialLeft(X,Y)
X accompanier Y	spatialLeft(X,Y), spatialSmaller(X,Y)
X depicts Y	spatialRight(X,Y
X annotates B	spatialBelow(X,Y), spatialAlign(X, Y)
X similar Y	spatialLeft(X,Y), spatialEqualSize(X,Y), spatialAlign(X,Y)
X component Y	spatialLeft(X,Y), spatialSmaller(X,Y)
X identifier B	spatialLeft(X,Y), spatialSmaller(Y,X
X precedes Y	temporalBefore(X, Y), spatialAlign(X, Y)

Although we have defined a single mapping from semantic relations to spatio-temporal relations in Table 3, the mapping from semantic to spatio-temporal relations is not always so straightforward. For example, when grouping media items, such as a group of works by the same artist, the actual number of media items, and their physical sizes are initially unknown. Unlike an HTML document, which is in principle, unbounded by page size, multimedia documents are generally less flexible.[7]. Although one approach to overcoming physical limitations is to display media items one after the other, this can be problematic. From a user's perspective, it is inadvisable to generate an over-repetitive presentation (e.g., slide show) which doesn't display new information or modify the layout. Moreover certain semantic relationships, such as *created*, might be better presented ordered spatially to illustrate the development in the works, rather than as a sequential slide show. In order to convey a grouping or ordering between media items, it would be better if a number of alternative spatiotemporal mappings were defined, in order of priority, if they cannot all physically fit on the screen at once. This problem is addressed by the *overflow stategies* of Rutledge et al. [19] and implemented

[7] Scrollbars besides and below the document allow a viewer to navigate to any point in a HTML document

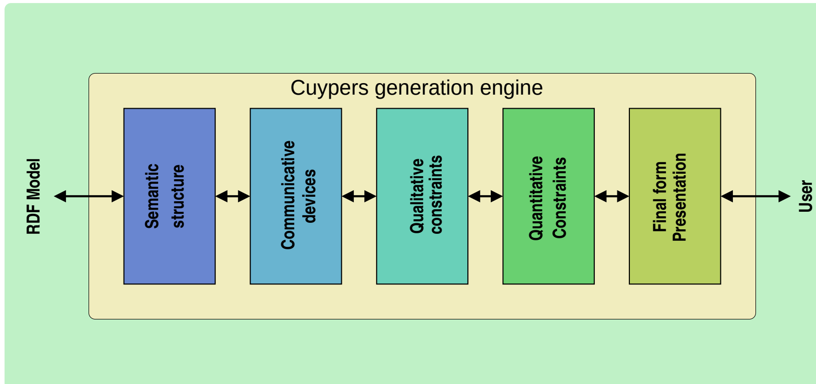


Fig. 4. Processing layers in Cuypers

within the Cuypers presentation engine, described in the next section. These strategies ensure that the intended semantics associated with the spatial-temporal relations, are respected and retained for an arbitrary number of media items of arbitrary sizes.

3.5 Cuypers Presentation Generator

The Cuypers presentation generation system [24] implements *overflow strategies* by applying constraint-solving mechanisms. It provides an abstract layer on top of the spatio-temporal relations found in multimedia document specification languages, such as SMIL [27]. Besides overflow strategies, the abstract layer allows Cuypers to adapt the presentation to other external factors such hardware constraints e.g., is the client using a multimedia PC or a mobile phone. Figure 4 illustrates the different processing layers within the Cuypers Presentation engine, which are described in more detail below.

1. **Semantic structure level** This level completely abstracts from the presentation's layout and hyperlink navigation structure and describes the presentation purely in terms of higher-level, 'semantic' relationships or rhetorical structure. This level within Cuypers is extended within our system by the semantic inferencing and semantic to spatio-temporal mapping processes described in the previous two sections which translate semantic relationships to the communicative devices of the next layer.
2. **Communicative device level** The highest level of abstraction describing the presentation's layout makes use of *communicative devices*[8] [19]. These are similar to the patterns of multimedia and hypermedia interface design described by [18] in that they describe the presentation in terms of well known spatial, temporal and hyperlink presentation constructs. An example of a communicative device is the *bookshelf*. This device can be effectively used in multimedia presentations to present a sequence of media items, especially when it is important to communicate the *order* of the media items in the sequence. How the bookshelf determines the precise layout of a given presentation in terms of lower level constraints can depend on a number of issues. For example, depending on the cultural background of the user, it may order a sequence of images from left to right, top to bottom or *vice versa*. Also its *overflow strategy*, that is, what to do if there are too many

[8] MFO's are called communicative devices in Cuypers.

images to fit on the screen, may depend on the preferences of the user and/or author of the document. It may decide to add a 'More info' hyperlink to the remaining content in HTML, alternatively, it could split the presentation up into multiple scenes that are sequentially scheduled over time in SMIL.

3. **Qualitative constraints level** An example of a qualitative constraint is "caption X is positioned below picture Y", and backtracking to produce alternatives might involve trying right or above, etc. Some final-form formats allow specification of the document on this level. In these cases, the Cuypers system only generates and solves the associated numeric constraints to check whether the presentation can be realized at all, it subsequently discards the solution of the constraint solver and uses the qualitative constraints directly to generate the final form output.
4. **Quantitative constraints level** To generate presentations of the same information using different document formats, we need to abstract from the final-form presentation. On this level of abstraction, the desired temporal and spatial layout of the presentation is specified by a set of format-independent constraints, from which the final-form layout can be derived automatically.
5. **Final-form presentation level** At the lowest level of abstraction, is the final-form presentation, which encodes the presentation in a document format that is readily playable by the end user's Web browser or media player. Examples of such formats include, HTML, SVG, and — the focus of our prototype — SMIL. This level is needed to make sure that the end-user's Web-client remains independent of the abstractions used internally in the Cuypers system, and to make sure that the end-user can use off-the-shelf Web clients to view the presentations generated by Cuypers.

3.6 User-directed Hypermedia Browsing

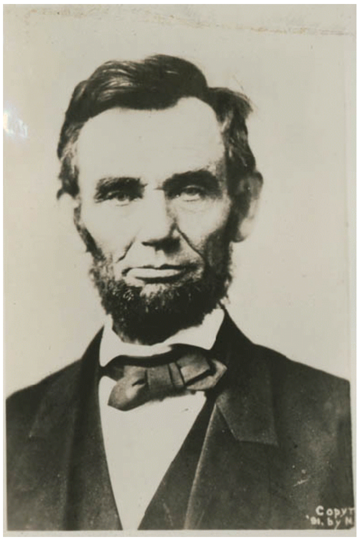
Currently the scope of a dynamically-generated presentation is limited to a single concept. A concept being, for example 'President Lincoln' or 'the Eiffel Tower'. However, each concept has a number of related concepts, such as speeches Lincoln has written or 'Gustav Eiffel' - these related concepts can be further explored by the user through hyperlinks from individual media objects within the current presentation.

Users are able to redirect a presentation's focus by clicking on particular individual media objects within the presentation to trigger a new iterative search process, using the metadata of the selected resource as the starting point. In this way, new presentations, focussing on a related concept can be dynamically generated. This step in the overall system is represented by the arrow from the presentation to OAI in figure 1.

4 Example

A user interested in the life of former American president *Abraham Lincoln* searches for the term 'Lincoln' across the OAI archives. The results include (apart from a certain amount of irrelevant material) a number of portraits, images of documents written by him, a news article reporting his assassination, an article about his life and images of the Lincoln Memorial. Two of the objects and their associated metadata are shown in Figure 5 below.

Because the 'subject' of the object on the left, is equivalent to the 'creator' of the object on the right, and the object on the left is of type 'image', it is possible to infer that the object on the left is a photograph of the creator of the object on the right. Further comparison of

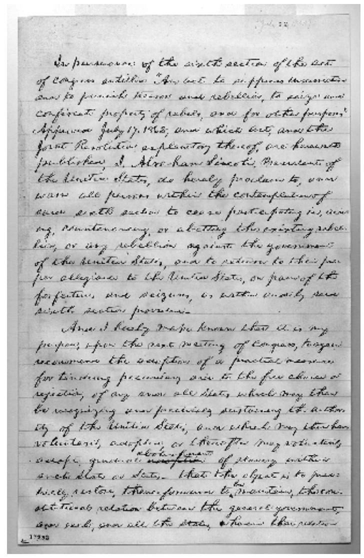


```
<title>Abraham Lincoln</title>

<creator>Gardner</creator>
<subject>Abraham Lincoln</subject>

<subject>President</subject>
<description>Photograph of
Abraham Lincoln taken in
Washington</description>

<date>1862</date>
<type>photograph</type>
<type>image</type>
```



In pursuance of the sixth section of the act of congress entitled "An act to suppress insurrection and to punish treason and rebellion, to seize and confiscate property of rebels, and for other purposes" Approved July 17. 1862, and which act, and the Joint Resolution explanatory thereof, are herewith published, I, Abraham Lincoln, President of the United States, do hereby proclaim to, and warn all persons within the contemplation of said sixth section to cease participating in, aiding, countenancing, or abetting the existing rebellion, or any rebellion against the government of the United States, and to return to their proper allegiance to the United States, on pain of the forfeitures and seizures, as within and by said sixth section provided.

And I hereby make known that it is my purpose, upon the next meeting of Congress, to again recommend the adoption of a practical measure for tendering pecuniary aid to the free choice or rejection, of any and all States which may then be recognizing and practically sustaining the authority of the United States, and which may then have voluntarily adopted, or thereafter may voluntarily adopt, gradual abolishment of slavery within such State or States. That the object is to practically restore, thenceforward to be maintain, the constitutional relation between the general government, and each, and all the states, wherein that relation

```
<title>Draft of the Emancipation
Proclamation</title>
<creator>Abraham Lincoln</creator>
<subject>Emancipation
Proclamation</subject>

<description>Draft of the
Emancipation Proclamation by
President Abraham Lincoln, July 22,
1862</description>
<date>22/7/1862</date>
<type>hand written document</type>
<type>image</type>
```

Fig. 5. Media items with associated metadata.

the dates associated with the objects indicates that the photo is of the creator at around the time the document was written.

Translating these inferred semantic relationships into spatial and temporal relationships generates a SMIL presentation which displays a chronological sequence of portraits on the left, with spatially beside them, playing in parallel, the display of the biography, images of documents he has created and concluding with the news paper article reporting his death. Figure 6 shows the user interface for the dynamically generated SMIL presentation.

5 Discussion, Conclusions and Future Work

5.1 OAI

The OAI repositories and the OAI protocol offered many advantages as a testbed for this work. In particular, the availability of interesting content and associated metadata, ease of use and the simplicity of the protocol will undoubtedly ensure the OAI's long-term sustainability

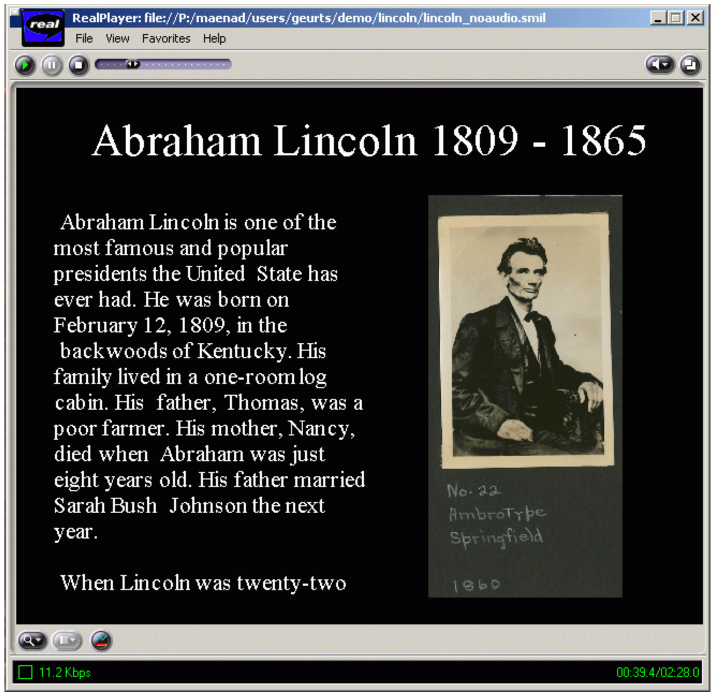


Fig. 6. Final form presentation

and potential for useful service provision. However a number of issues limited the OAI's effectiveness in the context of the work descibed here.

Firstly, it was difficult to find topics within the OAI archives, which are useful, interesting and have sufficient number and variety of related media objects to generate interesting presentations. Although existing OAI data providers make a relatively large number of items available, many of them are e-prints and hence of only limited suitability for creating multimedia presentations. Others, such as the Library of Congress, have multimedia but only on limited or generic topics. For example, pictures of a 'boat' or a 'park' are useless in this context without more specific metadata.

Secondly, the existing metadata is highly inconsistent, sometimes inaccurate and created with simple resource discovery in mind. Use of different thesauri and widely variable levels of detail and structure within the metadata make it difficult to effectively compare. Cole et. al. [8] encountered similar problems and suggest the use of metadata normalization for searching across OAI archives..

Finally, there is often either no link from the metadata record in OAI to the actual content or the link is broken. In many cases, the *dc.identifier* value points to the data provider's web page which may or may not contain the media item or a link to it. For automatic processing of metadata and generation of presentations, a direct link to the actual media item is essential.

5.2 Dublin Core

The unqualified Dublin Core metadata schema was chosen by OAI to maximize the chance of interoperablilty between data providers and because many of the participating institutions already have metadata in this format. It is also a simple yet flexible schema, suitable for many different subject areas. However we found that, in the context of the work described here, Dublin Core has some serious limitations.

Firstly unqualified Dublin Core is too simplistic to infer many rich or interesting semantic relationships. Disallowing qualifiers removes a level of detail from the metadata which is essential for anything other than fairly simplistic semantic inferencing. The extent to which this effects the quality of the final presentation is uncertain, since it is possible to produce a complete presentation with only a few relationships.

Secondly, the application of inferencing rules is further hampered either by unstructured metadata values or the use of incompatible schemas for many of the element values. For example, some data providers use textual values for the *dc.relation*, *dc.identifier* and *dc.source* elements, while others use URIs. This makes comparisons between values problematic. On the other hand, if particular controlled vocabularies or schemes are used for certain DC elements, it may not be applicable to migrate these values to other DC elements in order to infer semantically-related resources.

Thirdly, there is significant ambiguity over the purpose and content of many of the DC element's values. This is exacerbated by the use of unqualified Dublin Core. For example, in some cases the 'creator' is the creator of the digital surrogate. In other cases, the 'creator' is the person who created the object depicted in the digital surrogate. In some organisations, the 'creator' of a musical recording is the composer, while in others, the 'creator' is the primary musician. Likewise the *dc.date* value may be the date of creation of the original source item, the digital surrogate or the date of publishing in the OAI archive.

Finally, Dublin Core is not designed for describing multimedia resources. It is inadequate, in particular, for describing the fine-grained details such as segmentation, formatting and low-level audiovisual feature metadata which would be most useful for inferring interesting semantic relationships between multimedia objects.

None of these problems are new but their implications in the context of this work are accentuated and debilitating to the overall goal.

5.3 Conclusions and Future Work

Our first conclusion is that even given the limited range of multimedia resources available through OAI and the simple semantic relationships we have been able to infer from their metadata, we have been able to generate quite interesting and intelligent multimedia presentations. The advantage of our approach is that it can present previously unrecognized connections between related, distributed mixed-media resources in an easily interpreted, interesting and multi-layered display. Because the system is dynamic, new online resources with semantic relations to the search topic will be picked up and included automatically.

Our next conclusion is that much more interesting semantic inferencing would be possible if either qualified Dublin Core, or a richer model such as the ABC model [13] or MPEG-7 (where applicable), were used as the metadata model. The recent development of an MPEG-7 ontology [9], opens the way for richer semantic inferencing between resources described using MPEG-7 and resources described using other domain-specific ontologies.

However if we were to use more complex metadata schemas, such as ABC or MPEG-7, then we would need to replace the current iterative search and inferencing process with a more powerful and automated inferencing engine such as RDFGateway [11].

Although we have focussed on searches for information about 'people', specific semantic inferencing rules could be defined for searches on 'events', 'places' or 'physical objects'.

Expanding the types of searches and associated inference rules would expand the potential applications for the system e.g., automatic generation of multimedia news articles, obituaries, genealogies, museum presentations.

Our final conclusion is that in the next few years we will see the emergence of a new generation of search engines based on approaches similar to the one proposed in this paper. As the amount of multimedia on the internet expands, as the semantic web infrastructure develops and as more resources and ontologies are described using standards such as MPEG-7 and RDF, search engines are going to start delivering results in the form of automatically aggregated, knowledge-enhanced multimedia presentations.

6 Acknowledgements

The work described in this paper has been funded by the Cooperative Research Centre for Enterprise Distributed Systems Technology (DSTC) through the Australian Federal Government's CRC Programme (Department of Industry, Science and Resources). Thanks also to CWI, Lynda Hardman and Frank Nack for lending us Joost, paying his travel expenses and providing us with feedback and a copy of Cuypers for research purposes.

References

1. Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, May 2001.
2. Dublin Core Community. Dublin Core Element Set, Version 1.1. See http://www.dublincore.org/documents/dces/, 07/02 1999.
3. OAI Community. OAI Protocol for Metadata Harvesting 2.0. See http://www.openarchives.org/OAI/openarchivesprotocol.html.
4. OAI Community. Open Archives Initiative. See http://www.openarchives.org/.
5. OAI Community. Open Archives Initiative Registered Service Providers. See http://www.openarchives.org/service/listproviders.html.
6. T. Rist E. Andr. Generating Coherent Presentations Employing Textual and Visual Material. In *Artificial Intelligence Review, Special Volume on the Integration of Natural Language and Vision Processing, 9(2-3):147-165*, 1995.
7. Gregory Karvounarakis et. al. RQL: A Declarative Query Language for RDF. In *WWW2002 Conference*, Honolulu, May 2002.
8. T. Cole et.al. Now That We've Found the 'Hidden Web' What Can We Do With It? The Illinois Open Archives Initiative Metadata Harvesting Experience". In *Museums on the Web Conference 2002*, Boston, April 2002.
9. Jane Hunter. Adding Multimedia to the Semantic Web - Building an Mpeg-7 Ontology. In *International Semantic Web Working Symposium (SWWS)*, Stanford, July 2001.
10. USC Information Science Institute. Webscripter Project Overview. See http://www.isi.edu/webscripter/.
11. Intellidimension. Intellidimension's RDF Gateway. See http://www.intellidimension.com/RDFGateway/.
12. ISO IEC. *15938-5 FDIS Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Schemes*, March 2002.
13. Carl Lagoze and Jane Hunter. The ABC Ontology and Model. *Journal of Digital Information*, 2(2), 2001.
14. Alberto Reggiori Libby Miller, Andy Seaborne. Three Implementations of SquishQL, a Simple RDF Query Language. In *1st International Semantic Web Conference ISWC2002*, Sardinia, June 9-12 2002.
15. Library of Congress. Library of Congress: American Memory Collection. See http://memory.loc.gov/.
16. University of Illinois Library. University of Illinois Open Archives Collection. See http://bolder.grainger.uiuc.edu/uiLibOAIProvider/OAI.asp.
17. OpenVideo Organisation. OpenVideo. See http://www.open-video.org/oai/.
18. Gustavo Rossi, Daniel Schwabe, and Alejandra Garrido. Design Reuse in Hypermedia Applications Development. In *The Proceedings of the Eighth ACM Conference on Hypertext and Hypermedia*, pages 57–66, Southampton, UK, April 1997. ACM, ACM Press.

19. Lloyd Rutledge, Jim Davis, Jacco van Ossenbruggen, and Lynda Hardman. Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure. In *Proceedings of the International Conference on Multimedia Modeling 2000 (MMM00)*, pages 89–105, Nagano, Japan, November 13-15, 2000.
20. Michael Sintek and Stefan Decker. TRIPLE - A Query, Inference, and Transformation Language for the Semantic Web. In *1st International Semantic Web Conference ISWC2002*, Sardinia, June 9-12 2002.
21. Robert Tansley. *The Multimedia Thesaurus: Adding a Semantic Layer to Multimedia Information.* PhD thesis, Dept. of Electronics and Computer Science, Uni. of Southhampton, August 2000.
22. Douglas Tudhope and Daniel Cunliffe. Semantically Indexed Hypermedia: Linking Information Disciplines. *ACM Computing Surveys*, 31(4es), December 1999.
23. Frank van Harmelen, Peter F. Patel-Schneider, and Ian Horrocks. Reference description of the DAML+OIL (March 2001) ontology markup language. http://www.daml.org/2001/03/reference.html. Contributors: Tim Berners-Lee, Dan Brickley, Dan Connolly, Mike Dean, Stefan Decker, Pat Hayes, Jeff Heflin, Jim Hendler, Ora Lassila, Deb McGuinness, Lynn Andrea Stein, ...
24. Jacco van Ossenbruggen, Joost Geurts, Frank Cornelissen, Lloyd Rutledge, and Lynda Hardman. Towards Second and Third Generation Web-Based Multimedia. In *The Tenth International World Wide Web Conference*, pages 479–488, Hong Kong, May 1-5, 2001. IW3C2.
25. W3C. Semantic Web Activity. See http://www.w3.org/2001/sw/.
26. W3C. Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendations are available at http://www.w3.org/TR, February, 22, 1999. Editied by Ora Lassila and Ralph R. Swick.
27. W3C. Synchronized Multimedia Integration Language (SMIL 2.0) Specification. W3C Recommendations are available at http://www.w3.org/TR/, August 7, 2001. Edited by Aaron Cohen.
28. W3C. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Candidate Recommendations are available at http://www.w3.org/TR, 30 April 2002. Edited by Dan Brickley and R.V. Guha.