Content-based Querying Embedded In Multimedia Presentations

Doris Divotkey, Horst Eidenberger
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstraße 9-11, A-1040 Wien, Austria
{doris.divotkey, horst.eidenberger}@ims.tuwien.ac.at

***Abstract*—This paper presents a solution for integrating content-based multimedia retrieval with semantics-based content selection techniques, content adaptation, composition and delivery. The integration of several approaches of these domains provides users of digital information systems with a powerful toolkit. We focus on the generation of multimedia presentations and templates that can be reused. In this context, we propose a novel approach where content-based querying is embedded into multimedia presentations. Queries are executed when presentations are consumed. Hence, the requested media data are retrieved ad hoc.**

Keywords—visual information retrieval; content-based querying; multimedia presentations; user preferences and device capabilities

Topic area—Multimedia Databases (indexing and retrieval)

I. Introduction

The automatic selection of media for composing sophisticated multimedia presentations is a non-trivial task. If personal interests, preferences and skills have to be taken into account, it becomes even more complex. The focus of the paper is the application of content-based querying and multimedia retrieval techniques to enhance the composition of personalized multimedia presentations. In particular, we propose a novel approach for media selection by embedding content-based queries. Presentations contain encoded content-based queries to retrieve single media objects ad hoc during presentation consumption.

In case of multimedia data the semantic gap between high-level concepts and the low-level features that are used for similarity measurement means an additional problem. Modern content-based visual retrieval comes up with approaches to close or at least narrow this gap. Furthermore, the great number of existing hardware devices and network connections results in various multimedia formats. The goal is to provide the user with data appropriate for specific device capabilities. Therefore, adaptation in the sense of audio and video transcoding is an issue before the content is delivered to the user. The CoCoMA framework (Content and Context Aware Multimedia Content Retrieval, Delivery and Presentation) provides a solution for the generation of personalized and adapted multimedia presentations.

The paper is organized as follows: the next section gives an overview over the CoCoMA project. Section III describes more detailed the architecture of the CoCoMA application. Section IV focuses on the idea behind the embedded querying concept and outlines advantages as well as drawbacks. Section V shows use case scenarios.

II. Background: The CoCoMA Project

The CoCoMA project is an activity within the DELOS Network of Excellence on Digital Libraries. The goal is the combination of content-based techniques and the usage of semantic media annotations for the selection of multimedia content and the creation of personalized presentations.

The CoCoMA framework comprises a set of user interfaces for specification of hardware settings, preferences, media annotations and queries on the one hand and various visualization tools on the other hand. A sophisticated framework is provided for selection and composition of content on the basis of user preferences. For that purpose, the user profile is compared with the annotated media metadata. The adaptive media store provides the media and adapts it according to the user's hardware and software requirements. Due to clearly defined interfaces the adaptation engine can be easily extended by video filters and converters. The composition component inserts the adapted media into a tailor-made presentation for the user.

The CoCoMA framework components are implemented in Java. Web services are used for communication. User settings are expressed in XML based on standards like W3C CC/PP and *Usage Environment* of the MPEG-21 DIA (Digital Item Adaptation) architecture [6]. Media annotations follow a core ontology (based on MPEG-7 MDS [5]) and domain-specific extensions both based on RDF/OWL. [7]

The software toolkit comprises the implementation of the framework in Java and the MPEG-7 and MPEG-21 content descriptors. Furthermore it contains XML representations of the core ontology and domain-specific extensions for domains like soccer and news. The entire toolkit will be released as open source.

III. Architecture

The main components of the CoCoMA architecture and their interaction are shown in Figure 1. The framework consists of the following parts:

- The *Content Selection, Composition and Presentation Framework* is the core component responsible for the

analysis of user settings, the choice of appropriate content and the composition of individual presentations of the retrieved and adapted media. Thereby, constraints given by the Content Authoring System can be incorporated. [2]

- The *Adaptive Media Store* is the point for media and metadata access. Both, semantics-based (MPEG-7 MDS, MPEG-21 DID) and content-based descriptions (MPEG-7 audio and visual descriptors) are supported. [4]
- The *Querying Component* uses semantics- and content-based retrieval techniques to identify relevant media objects according to the defined querying parameters. [3]
- The *Multimedia Content Authoring System* allows for the definition of constraints for temporal and spatial relations between media objects. These constraints are used for the creation of presentations. [1]
- The *Annotation Component including Semantics-based Metadata Extraction* allows for the declaration of textual annotations. Media annotations are used for retrieval by keyword on the one hand and as check point to determine an adaptation plan according to the user's preferences and device settings on the other hand.
- The *Content-based Metadata Extraction Component* extracts content-based features. Content-based descriptors can be either applied for retrieval or adaptation tasks. For media objects stored in the media store the extraction of descriptions can be performed in advance. Additionally, if the query component asks for media descriptions not yet available (e.g. for live streamed media), they can be extracted on the spot.

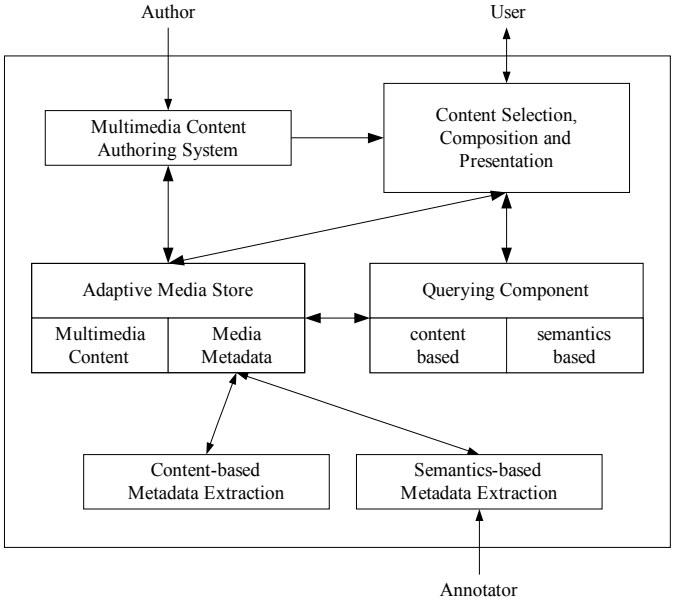

Figure 1 CoCoMA components and their interaction

IV. The Embedded Querying Concept

In this subsection we present a novel approach to apply content-based querying within the presentation step of a multimedia composition and delivery application. Moreover, we point out advantages and drawbacks of the concept.

A. The Concept

The commonly applied process for the generation of multimedia presentations is given as follows:

The user defines his area of interest, his preferences in terms of media types etc. as well as device and networking capabilities. Due to these settings, the framework determines the best fitting presentation template. The system selects the individual media objects and composes a personalized presentation adjusted to the user's requirements. For identifying relevant media for the user, content-based querying techniques are used. As shown, this straightforward approach applies querying first and builds up presentations on the bases of the retrieved results afterwards. Despite its advantages, one has to consider a few drawbacks as well:

- The media have to be available during query execution. This means they have to exist already during presentation generation. Therefore, live streaming can not be handled.
- The presentations contain hard-coded links to media. Exchanging media objects entails changes in the presentations. This may be time-consuming and error-prone.

What we propose is to change the order of the presentation building step and the querying step. Hence, we build a presentation first and delay the querying to the actual execution of the presentation. This involves changes in the presentations.

Such multimedia presentations do not only contain direct references to media objects. Additionally, the selection of content is encoded within the presentation. In other words, the composition component introduces references to media objects that are not yet known. These place holders are therefore references to a content-based retrieval connector which has the task to fill in a matching media object. Query parameters specify the requested media object and have to be encoded in the query link. In this case, the composition component defines appropriate parameters according to the user's information demand (taking preferences and hardware settings into account) and embeds the query and the parameters into the presentation.

Besides, it is assumed that the query connector will retrieve a single, suitable media object according to the specified query parameters. This is performed ad hoc when the presentation is consumed.

B. Implementation

Figure 2 shows a presentation that consists of a collection of building blocks arranged due to temporal and spatial constraints. A building block itself may contain several media objects of different type also organized by structural, spatial and temporal relations. An existing building block can be reused in other presentations as well. Thereby, both simple and complex – more sophisticated – presentations can be created fast and easily.

Before delivery, a presentation has to be transformed to a multimedia presentation format the user's client application can handle. This is done according to the user's preferences and device capabilities (e.g. plain HTML or SMIL).

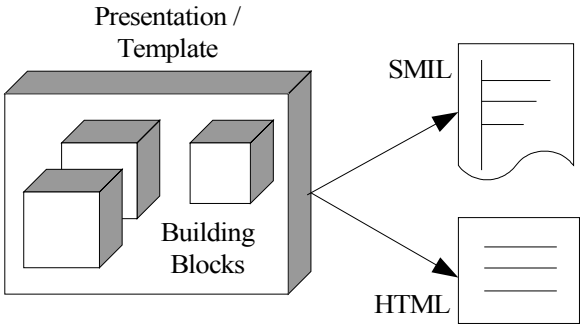

Figure 2 Presentation with building blocks

Figure 3 shows a schematic drawing of the usage of content-based querying on the level of presentations. The individual media objects are referenced by URIs. The client application (e.g. a SMIL player) requests the individual media objects of the presentation via the specified URI. An URI need not be a direct reference to a media object, but can as well be a reference to a querying component that returns a media object. The composed presentations contain therefore links to media objects retrieved ad hoc by content-based queries during presentation execution. There is no difference for the player whether requesting a media object directly or through the content-based retrieval connector to which the reference in the presentation links. The CBR connector returns a single media object fulfilling the query specification.

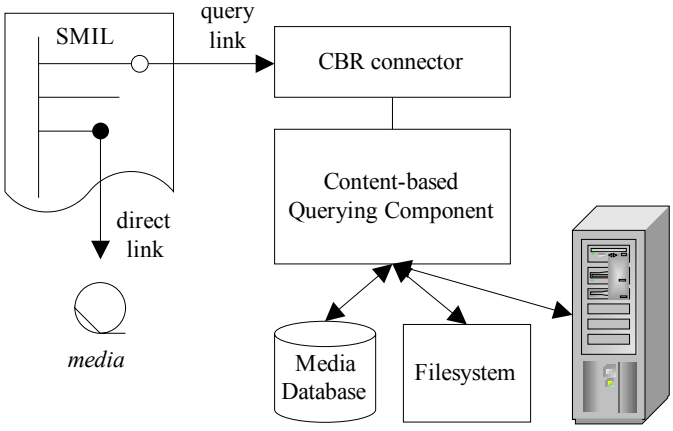

Figure 3 Embedded content-based querying

The following example shows a simple SMIL presentation that contains both a direct reference to an image file and a link embedding a content-based query for retrieving a text.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<smil xmlns=
      "http://www.w3.org/SMIL20/Language">
<body>
  <par>
    <img src="http://www.ims.tuwien.ac.at/
lion.jpg"/>
    <text src="http://www.ims.tuwien.ac.at/
cbrConn?type=text&keywords=lion+wildlife
+habitat"
          alt="Description about the wildlife
habitat of lions"/>
  </par>
</body>
</smil>
```

The presentation shows the image of a lion and a description about the wildlife habitat of lions whereby the <*text*>-tag contains an embedded query. The CBR connector (*cbrConn*) initiates a query by keyword as the *type* is set to *text*. The parameters (*keywords*) are passed on to the actual text retrieval component that determines and returns the best matching document. In addition, the connector can include externally-stored user profiles in the querying process. For the example above the language of the searched text may be defined as user preference.

The query parameters encoded in the URI depend on the used querying component and model. The approach is not limited to text retrieval. It qualifies in particular for image and video retrieval.

C. Advantages and Drawbacks

Decoupling the querying from the generation of multimedia presentations by postponing the querying task to the presentation execution provides independence of the generation of presentations and the actual content. A major advantage of this approach is that there are no assumptions about the media collections used. Neither the kind of storage (database, file system, web, ...) nor the structure is restricted. The media objects can be stored in distributed media collections and their organization can be arbitrary.

Furthermore, the desired media object need not be existent when generating the presentation. Hence, presentations can be produced offline. Another obvious benefit is the reusability of such multimedia presentations with embedded queries. One and the same presentation can be applied to different media collections. The presentation adapts automatically according to the changed data set. The use case scenarios in the next section provide more details and application domains of the presented technique.

There are a few drawbacks to mention as well. As the querying is done during presentation execution, it has to be performed in real-time. It is difficult to guarantee this requirement as several factors have negative influence in this respective: the complexity of the used features, whether the descriptions are already available or have to be extracted on the spot and the size of the media collection. Moreover, the query functionality is limited as the query parameters are encoded in the URI to the query connector and therefore the amount of information that can be encoded is restricted.

Considering this, the presented approach is not meant as a replacement but a reasonable enhancement for sophisticated multimedia presentations. Presentations with embedded

content-based queries are templates for real presentations. In case the presentation generation process is very time-consuming due to extensive querying, the retrieved presentation may be stored and reused to avoid recurrent computation and facilitate real-time execution of the actual presentation.

V. Use Case Scenarios

The following subsections show the usefulness of the presented approach in practice by means of three use cases.

A. On demand summaries of live sport transmissions

Among other things, research in the domain of sport events deals with the analysis of live streamed video and the automatic extraction of semantic information. In this regard, it would be convenient to have a system that offers on demand a summary of recently happened, important events (e.g. goals, fouls, ...). With the proposed approach such a system can be simply provided by creating a template for a presentation containing the designated information. An appropriate template for this scenario could consist of the current score and a replay of the last goal.

With such a predefined setting the actual presentation adjusts automatically according to the progress within a match. Every time the user requests an update of the presentation the content-based querying component queries for the scene showing the recently scored goal and determines the current score.

The described scenario could not have been achieved with conventional methods because on the one hand the content that should be presented is in generation itself and therefore not present yet. On the other hand even after transmission of the match the content is not available in an appropriate form, as the video stream has to be segmented first. Determining the "recent goal" means semantics- and content-based filtering of the video stream and retrieving the specific segment of the stream fulfilling the query.

B. Tourist Information System

The creation of attractive multimedia presentations is a time-consuming and costly task. A tourist information system provides both content that is constant for a long time (e.g. a documentary about a church) and rather frequently changing information like current exhibitions and events.

Embedded content-based queries in tourist information presentations are especially suitable for varying content. In contrast to database queries, this approach has the advantage that it is independent from the underlying data storage scheme. The structure of media collections need not be considered when presentations are created. Moreover, the data organization can be changed afterwards without having to adapt the presentations. Beyond, the media objects need not be indexed or classified at all.

Such presentations can be used as templates to build new presentations efficiently. Once created they can be applied on different media collections or with different context. When the execution of a presentation is repeated, the queried content changes according to the changed situation (e.g. current events are updated automatically).

C. Web-based Presentations

Although the world wide web brings considerable advantages, the access to media like images and video is rather limited. By now, the common approach is querying by keyword which is quite inefficient for images and video. Content-based retrieval systems try to come up with better solutions. Unfortunately, the usage of these systems requires profound knowledge of the user in terms of selection of features, similarity measures and so on. Hence, content-based querying is not yet applied on a large scale.

The presented approach provides every user with the possibility to reap the benefits of content-based querying without having to be a professional. Content authors and experts in content-based querying techniques create presentation templates containing embedded content-based queries. Users may choose among these templates and in combination with their preferences, they query the web and gather their personalized presentation simply by executing the predefined template.

VI. Conclusion and Future Work

The paper presents a novel approach where content-based querying is embedded in multimedia presentations to create reusable templates and provide personalization according to user preferences. Since the query step happens during the execution of the presentation, the media need not be existent when the presentation is created.

Future work will comprise the implementation of the proposed approach and its integration into the presented CoCoMA framework. Thereby, an appropriate querying language for visual information retrieval has to be defined.

VII. Acknowledgment

The CoCoMA project is an activity within the DELOS Network of Excellence on Digital Libraries funded by the 6th Framework Programme of the European Union.

References

[1] E. Bertino, E. Ferrari, A. Perego, D. Santi, "Constraint-based Techniques for Personalized Multimedia Presentation Authoring," *AVIVDiLib Workshop*, 2005

[2] S. Boll, "MM4U – A Framework for creating personalized multimedia content," *Proc. International Conference on Distributed Multimedia Systems*, September 2003

[3] H. Eidenberger, C.Breiteneder, "VizIR – A Framework for Visual Information Retrieval," *Journal of Visual Languages and Computing*, vol. 14, pp. 443-469, 2003

[4] K. Leopold, D. Jannach, H. Hellwagner, "Knowledge-based Media Adaptation," *SPIE Proceedings Series*, vol. 5601, pp. 111-120, October 2004

[5] P. Salembier, J. R. Smith, "MPEG-7 Multimedia Description Schemes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, June 2001

[6] J. R. Smith, "MPEG-21 Digital Item Adaptation: Enabling Universal Multimedia Access," *IEEE Multimedia*, vol. 11, pp. 84-87, Jan-Mar 2004

[7] C. Tsinaraki, P. Polydoros, S. Christodoulakis, "Interoperability support for Ontology-based Video Retrieval Applications," *Special Issue on the IEE VIS Journal*, CIVR 2004