SEARCHING ANNOTATED BROADCAST CONTENT ON MOBILE AND STATIONARY DEVICES

Gabriella Kazai[1], Mounia Lalmas[1], Alan Pearmain[2] and Marie-Luce Bourguet[1]
Department of Computer Science[1] and Department of Electronic Engineering[2], Queen Mary University of London
{gabs, mounia, mlb}@dcs.qmul.ac.uk[1] and Alan.Pearmain@elec.qmul.ac.uk[2]

ABSTRACT

This paper describes, within the framework of the European IST SAVANT project, the development of a search engine that supports transparent access to annotated broadcast content from different types of user devices. For this purpose, the search engine makes use of emerging standards such as MPEG-7, TV-Anytime and MPEG-21.

KEYWORDS

MPEG-7, TV-Anytime, MPEG-21, broadcast content, retrieval, filtering, recommendation

1. INTRODUCTION

The growth of digital TV and increasing user mobility are the main driving forces behind the European IST SAVANT (Synchronised and Scalable AV content Across NeTworks) project (www.ist-savant.org). The project is developing broadcast technologies that allow end-users to access interrelated and synchronised multimedia content in an intelligent and transparent manner on stationary and mobile terminals and under varying network conditions. Intelligent and transparent access to broadcast content requires content to adapt to the various terminal and network characteristics and user interests. In SAVANT, this adaptation relies on semantics-based annotation of the content, that is, metadata associated with the multimedia content. The employed metadata structures are based on established and emerging standards, such as MPEG-7, TV-Anytime and MPEG-21. This paper describes the search engine component of the SAVANT system, which builds on this metadata and allows end-users to obtain multimedia content of interest to them using different user terminals.

2. THE SAVANT SYSTEM

The overall architecture of the SAVANT system consists of three main components: *content creation and annotation*, *delivery* and *content access* systems. The broadcasting chain starts with the content creation and annotation system whose output is the MPEG-2 main content and additional content (e.g. MPEG-4, HTML, JPEG, etc.) annotated with metadata. This output is fed to the content delivery system, which is responsible for the smart routing of content via multiple transmission channels (e.g. DVB and IP) to the content access system. The content access system includes a combination of the following terminals: (1) A Set Top Box (STB) with a TV set, which forms the core of the overall system providing large storage and also used as a gateway server; (2) a Tablet PC operated via touch screen and keyboard, which is used as a portable device at home; and (3) a PDA, which acts as the mobile device for “on the move” usage. Each of these terminals can be used by itself or in combination with others. For example, the STB, used simultaneously with several PDAs allows for group viewing of the main broadcast content while supporting personalised viewing of additional content.

The content access system includes the following sub-systems. The *content presentation* sub-system controls and synchronises the different players for the various multimedia content (e.g. MPEG-2, MPEG-4, HTML) to be displayed on the various user terminals and handles user requests for additional content. The

content and service adaptation sub-system supports content (transcoding, summarizing and abstracting) and metadata adaptation. The *storage manager* sub-system manages the different remote (DVB and Telecom) and local channels, and serves as the basic function unit for all types of content managed by the terminal (e.g. broadcast content extracted from DVB or IP channels, metadata, user profiles, and terminal characteristics). The *search engine* sub-system is responsible for the access to the main and additional multimedia content. It includes a *user manager* sub-component for the maintenance of user profiles, and supports three main functions: retrieval, filtering and recommendation. Retrieval allows users to search the content of the STB's hard disk (local storage) for information relevant to the currently viewed programme (or programme item) or regarding an arbitrary information need expressed as a query. Filtering provides message alerts to users based on their interests by monitoring the incoming broadcast content. Recommendations, both at programme and programme item level, are provided to users browsing content listings or archives, where the list of available content items (e.g. list of news items in a news bulletin or programmes in an Electronic Programme Guide (EPG)) are re-ordered according to the user's profile so that items of interest appear at the top of the list.

3. EMPLOYED METADATA STANDARDS

The development of transparent and intelligent access to the multimedia content requires this content to be semantically annotated with appropriate metadata. As a result of previous studies (Kazai et al, 2003, Liu et al, 2002), MPEG-7 (2001), TV-Anytime (2002) and MPEG-21 (2002) were selected to provide the semantics-based annotation of the multimedia content. The developed metadata structures combine elements from all three standards.

```
<Mpeg7Main>
<ContentDescription xsi:type="ContentEntityType">
 <AudioVisualContent xsi :type="AudioVisualType">
  <AudioVisual id="AUExample">
   <MediaInformation>…</MediaInformation>
   <CreationInformation>
    <Creation> <Title>Eating out in Budapest</Title>
    <Abstract>A documentary visiting famous restaurants in
              Hungary's capital</Abstract></Creation>
   <Classification><Genre>Documentary</Genre></Classification>
   <RelatedMaterial>…</Related Material> </CreationInformation>
   <SegmentDecomposition decompositionType="temporal">
    <Segment id = "S1_Matthias Cellar">
     <TextAnnotation>
      <FreeTextAnnotation>Opening scene with celebration dinner at
          the famous Matthias Cellar restaurant </FreeTextAnnotation>
      <StructuredAnnotation> <Where>Matthias Cellar </Where>
      </StructuredAnnotation>
      <KeywordAnnotation>
       <Confidence>0.75</Confidence>
       <Keyword>Matthias Cellar</Keyword>
      </KeywordAnnotation>
     </TextAnnotation>  </Segment> …
</Mpeg7Main>
```

Figure 1. Example of a MPEG-7 description

```
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS">
 <Item id="MPEG21-Example">
  <Choice minSelection="1" maxSelection="1">
   <Descriptor> <Statement type="text/text">Content type
         </Statement> </Descritor>
   <Selection select_id="MPEG-2"> <Descriptor> <Statement
       type="text/text"> MPEG-2 </Statement> </Descriptor>
   </Selection>
   <Selection select_id="MPEG-4"> <Descriptor> <Statement
       type="text/text"> MPEG-4 </Statement> </Descriptor>
   </Selection>
  </Choice>
  <Item> <Condition require="MPEG-2">
   <Component> <Resource ref="example.mp2"
       type="audiovisual/mp2"> </Component>
   <Component> <Resource ref="example.mp7"
       type="text/mp7"> </Component>
  </Item>
  <Item> <Condition require="MPEG-4">
   <Component> <Resource ref="example.mp4"
       type="video/mp4"> </Component>
   <Component> <Resource ref="other.mp7"
       type="text/mp7"> </Component>
  </Item> … </DIDL>
```

Figure 2. Example of a MPEG-21 description

MPEG-7, which provides a rich set of description tools, is used to describe the structure and semantics of the multimedia content. The MPEG-7 file, shown in Figure 1, describes a video decomposed using temporal segmentation. We use semantically rich, high-level descriptor schemes (DS) e.g. CreationInformation and for text annotation, the FreeTextAnnotation, StructuredAnnotation and KeywordAnnotation DSs, which provide semantic descriptions both in natural language and in structured form (e.g. Who, WhatAction, Where, etc.).

TV-Anytime, which is a metadata standard developed to define specifications for programme level content descriptions, is used to allow viewers to find, navigate and manage content from a variety of sources including enhanced broadcast, Internet and local storage. The metadata includes attractors (e.g. title, synopsis, genre, cast and awards) to aid the acquisition of available content organised, for example, in EPGs.

Finally, MPEG-21 is used to describe the annotated content as a collection of Digital Items, thus enabling the transparent delivery of the content across different networks and to different user devices. MPEG-21 was

found most suited for this purpose as it is a metadata framework that aims to enable transparent access to multimedia resources across a wide range of networks and devices. MPEG-21 targets the adaptation of Digital Items defined as structured digital objects with standard identifications and descriptions. It describes a variety of dimensions, including user preferences (e.g. display, accessibility and mobility characteristics), terminal capabilities (e.g. hardware and software properties, device profiles indicating the supported media formats, e.g. MPEG-2), networks (e.g. delay, error and bandwidth) and delivery (supported transport protocols, e.g. TCP/IP, and the types of connections, e.g. multicast). Figure 2 describes a Digital Item composed of two sub-items: an MPEG-2 content described by the MPEG-7 metadata shown in Figure 1, and an MPEG-4 video also associated with some (or even the same) MPEG-7 description.

4. THE SEARCH ENGINE

The SAVANT search engine is responsible for the transparent and intelligent retrieval, filtering and recommendation of multimedia content arriving via different delivery channels (DVB, IP). For this purpose, the search engine implements functions to index the metadata describing the broadcast content (i.e. MPEG-7 and TV-Anytime), formulate and process user queries, and provide relevance-based ranking.

4.1. Indexing the Metadata

The search engine indexes the MPEG-7 metadata that describes the multimedia content at the media stream level and the TV-Anytime metadata, which provides programme level description. During the indexing process functions such as stop-word (e.g. “the”, “an” etc.) removal and stemming are applied. Content analysis and statistical techniques are then used to extract and weight keywords and classifications from the metadata. Weighted keywords allow searching with respect to content (e.g. usually with respect to data contained in the FreeTextAnnotation descriptors), and classifications allow searching attributes (e.g. genre as documentary). TV-Anytime metadata is used to enable retrieval, filtering, and recommendation of content from the programme listings of EPGs. This is based on description and classification information extracted from the BasicContentDesription and BasicSegmentDescription elements of the ProgramInformation DS.

The metadata is indexed off-line for content stored in local storage and on the fly for live broadcast using the HySpirit retrieval platform (Rölleke et al, 2001). The indexed metadata is represented using probabilistic relational algebra, where the textual values of the descriptors, extracted during content analysis, are stored as *content* and the nested retrieval units (e.g. Segment, StillRegion) are referenced as *contexts*. The storage of the structural relationships between contexts enables the propagation of content values along the MPEG-7 and TV-Anytime metadata tree. This allows the retrieval process to return multimedia content elements of varying granularity, e.g. a StillRegion, a Segment or a whole video depending on which unit contains relevant information for the user (Pearmain et al, 2002).

4.2. Query processing

Retrieval is initiated with a user query, which is either automatically derived from the metadata description of the currently viewed content or manually produced by the user with the help of an input device. With the exception of the STB, all SAVANT terminals are equipped with a keyboard or pen. For the STB, users are traditionally limited to the use of a remote control, which, although suitable for browsing, is cumbersome for query input. As an alternative, a PDA can be used as an additional input device to the STB. The filtering and recommendation components use user profiles as queries, which describe the user’s topical interests. These can be manually built through a dialogue with the user, or can be automatically created by monitoring the actions of a user over a period of time. By examining the type of content that he/she chooses to view and by employing learning algorithms a user profile can be constructed and then continually updated to reflect the changing interests of the user. Automatically derived queries are represented in MPEG-7 and TV-Anytime syntax, while manually entered queries can be in natural language or in XPath dialect (Liu et al, 2002), thus allowing users to issue searches requesting, for example, films where a certain actor is cast. Once the query is formalised it is transformed, using a process similar to that used for metadata indexing, to a suitable format and passed to the ranking engine along with the MPEG-21 metadata describing the capabilities of the requesting terminal.

4.3. Relevance-based ranking

The relevance-based ranking of multimedia content is based on a matching function that provides a ranking of items according to their estimated relevance to the user's query, taking into account the capabilities of the requesting user terminal. Matching is based on a probabilistic framework, where the degree of relevance is estimated via statistical models derived from the indexed metadata and the processed query. The matching function, implemented using HySpirit (Rölleke et al, 2001), is based on the combination of feature-based *content* retrieval and *context* matching using path algebra. Only those resources are considered whose media type is appropriate for the capabilities of the user device described using the MPEG-21 TerminalCapabilities DS. For example, if the user query is "Restaurants in Budapest" and the terminal is capable of playing MPEG-4 but not MPEG-2, then given the Digital Item in Figure 2, only the MPEG-4 component would be retrieved with a calculated probability of relevance based on its MPEG-7 description. Retrieval can also be performed remotely, by passing the query onto external search services. For example, in the case of a PDA, the actual search and retrieval takes place in the STB. Filtering and recommendation use the same matching function, the difference being that queries are largely static user profiles, which are then compared to incoming content indexed on the fly. Filtering and recommendation also makes use of a threshold value, and only content items with estimated relevance degree higher than the threshold are forwarded to the user.

5. CONCLUSIONS

The recommendation component of the search engine has been demonstrated as part of the first prototype of the SAVANT system at the Berlin International Broadcast Fair (IFA 2003). The recommendation system indexed the metadata describing the broadcast content and evaluated it against the current user's profile producing a ranking of content items, where the most interesting items were then recommended to the user on the current user terminal (e.g. TV, PDA, Tablet PC). Interests within a user profile were implemented as a list of likes and dislikes with probability values reflecting degrees of interests. For IFA, two predefined user profiles were used, proposed by our broadcast partner (Rundfunk Berlin-Brandenburg) based on realistic user expectations in the scenario adopted for IFA, which consisted of news programmes broadcast on a given day.

Our future aims are to further develop the content access system, the final version of which is to be demonstrated at the IBC broadcast fair (September 2004, Amsterdam). For IBC, instead of simulating user profile construction, we aim to implement learning algorithms that will create and adapt user profiles following the changing interests of users. A sport scenario is also being developed, which will be used in conjunction with the news scenario developed for IFA. We will continue to base our search engine on the use of the MPEG-7, TV-Anytime and MPEG-21 metadata standards as they support the implementation of transparent and intelligent access to multimedia content in broadcast applications.

ACKNOWLEDGEMENT

SAVANT is funded by the European Union under the contract number IST-2001-34814.

REFERENCES

ISO MPEG-7, Part 5 - Multimedia Description Schemes, ISO/IEC JTC1/SC29/WG11/N4242, 2001.

ISO MPEG-21, Part 7 - Digital Item Adaptation, ISO/IEC JTC1/SC29/WG11/N5231, 2002.

Kazai, G. et al, 2003. Using Metadata to Provide Scalable Broadcast and Internet Content and Services, *WIAMIS Workshop*, London, pp 487-492.

Liu, P. and Hsu, L.H., 2002. Queries of Digital Content Descriptions in MPEG-7 and MPEG-21 XML Documents. *XML Europe*, Barcelona.

Pearmain, A. et al, 2002. Using MPEG7 at the Consumer Terminal in Broadcasting, *EURASIP Journal on Applied Signal Processing*, Vol. 4, pp 354-361.

Rölleke, T. et al, 2001. The HySpirit Retrieval Platform. *ACM SIGIR Demonstration*, New Orléans.

TV-Anytime Provisional Specification 1.3, http://www.tv-anytime.org, 2002.