Retrieval of MPEG-7 based Semantic Descriptions

Mathias Lux[1], Michael Granitzer[2]

Institute for Knowledge Management and Visualization
Technical University of Graz
Inffeldgasse 21a
8010 Graz, Austria
mathias.lux@tugraz.at

Know-Center Graz
Competence Centre for Knowledge Based Applications R&D
Inffeldgasse 21a
8010 Graz, Austria
mgrani@know-center.at

Abstract: With the growing amount of people using the Internet, and creating digital content and information, knowledge retrieval becomes a critical task. Knowledge retrieval allows people to find content they search for when they need it. Ongoing efforts provide frameworks and standards for annotating digital and non-digital content semantically to describe resources more precisely and processable in comparison to simple descriptive structured and unstructured metadata. One example for a framework for semantic annotations is MPEG-7. But although the MPEG group provides with MPEG-7 a useful and well defined theoretical framework for the creation of semantic annotation, retrieval of the annotations is not discussed. In this paper the authors present a retrieval process for MPEG-7 based semantic annotations founded on well proved information retrieval techniques, namely query expansion and regular expressions. Additionally a prototype implementation extending an existing MPEG-7 retrieval application will be presented.

1 Introduction

In traditional libraries metadata plays a central role, as keywords and taxonomies provide short and meaningful descriptions and cataloguing. It provided for a long time the only alternative way to inspecting all available books of finding what users need within the inventory of a traditional library. In digital libraries this context was “digitized” but remained quite similar to the original concept. Current trends show that the efforts and achievements of the information retrieval research area are integrated to enhance digital libraries [Lossau2004], [Summan2004]. On the other hand much of the metadata based methods of digital libraries have been adopted in knowledge management, knowledge discovery and information retrieval (i) for providing an application area for techniques like metadata extraction and automatic taxonomy creation and (ii) for enhancing knowledge management, discovery and retrieval by using metadata based retrieval techniques. Especially in the latter field techniques based on query expansion using thesauri or ontologies are very successful. Multimedia retrieval heavily depends on such techniques and the appropriate metadata. Content based image and video retrieval requires the preprocessing and indexing of content before query time, this pre-processing is the extraction of low level metadata. An often discussed topic in content based image retrieval is the *semantic gap* ([DelBimbo1999], [Smeulders2000]), which defines the difference between automatically extracted image features and the understanding or description of visual information of a user. If the semantic gap can be bridged by retrieval mechanisms no annotation would be necessary.

Right now semantic descriptions have to be created, at least in parts, manually. Human computer interaction (HCI) methods and information retrieval methods exist, that support the user in the annotation task. Different formats and approaches for the storage and definition of semantic descriptions are currently discussed and in use, wherefrom MPEG-7 is one of them.

The field of information retrieval offers lot of already well proved methods and techniques, which are known to work fast and reliable, especially in the field of text retrieval. Many research prototypes, text retrieval tools and search engines, which often represent current academic developments, are freely available. That is why researchers can use existing basic text retrieval methods easily to create a more complex framework, for example for retrieval of semantic descriptions based on text retrieval methods, concentrating only on new parts.

1.2 Related Work

Many application oriented research projects use MPEG-7 as their base standard and extend it to support their application domain or pick out single applicable parts. A very prominent research project is Marvel [IBM2004a]. Marvel is an MPEG-7 based video retrieval system which can extract up to 200 different semantic concepts like “summer”, “winter”, “car” or “aeroplane”, from video streams automatically. The Marvel focuses on the extraction of high level metadata, but it extracts only on stand alone concepts and no relations between the single concepts. IBM already demonstrated their MPEG-7 knowledge and commitment in the research project VideoAnnex [IBM2004b], which allows the manual annotation of video segments after automatic segmentation of the video streams. QBIC [Flickner1995] is a framework for content based image retrieval and shows IBM’s abilities in this area. The Informedia project [Wactlar2002] extracts metadata from video and film archives for e.g. retrieval based on this metadata and visual video summaries. Marvel and Informedia try to bridge the semantic gap by content analysis. The retrieval of semantic descriptions is not discussed before semantic descriptions can be extracted automatically.

Other projects already use semantic metadata. For the creation of semantic metadata manual and semiautomatic annotation is used instead of extraction. An MPEG-7 driven framework for managing semantic metadata for audiovisual content was presented in [Tsinaraki2003]. The annotation is based on simple forms and on a fixed domain ontology, the retrieval is restricted to querying metadata for specific temporal video segments, which is visualized using a transformation from MPEG-7 to TV-Anytime metadata. The IMB (short for the German words *Intelligente Multimedia-Bibliothek*) supports annotation and retrieval of semantic descriptions with a visual annotation and query creation tool [Mayer2004]. Annotations in IMB are based on one specific domain ontology and the retrieval mechanism supports only exact matches based on existing instances of semantic objects, not supporting any ranking or fuzzy matching. Both of these research projects only use single domain ontologies, restricting the possibilities of the MPEG-7 Semantic DS. Furthermore the retrieval mechanisms are pure data retrieval mechanisms which do not support partial match or relevance calculation along with common information and knowledge retrieval features like relevance feedback or user models.

2 MPEG-7 based Semantic Descriptions

MPEG-7 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the MPEG-1 and MPEG-2, and the MPEG-4 standards, which are mainly video coding and delivery standards [Martinez2003]. A quite new and not yet finished effort is the specification of MPEG-21. It aims at defining a normative open framework for multimedia delivery and consumption for use by all the participants in the delivery and consumption chain, including for example intellectual property management, digital rights management and content coding and delivery [Bormans2002].

MPEG-7, which is formally named the “Multimedia Content Description Interface”, is a standard for describing the multimedia content, which can be passed onto, or accessed by, a device or a computer code. MPEG-7 is not aimed at one specific application or domain of applications but tries to include as many different aspects of different application domains as possible. In a first step MPEG-7 defines a XML-Schema based definition language, called Description Definition Language (DDL). Using DDL the MPEG specified Descriptors (D), which in general represent a metadata element, e.g. a visual feature or a textual description, and Descriptor Schemes (DS), which specify the structure and semantics of Descriptors. For a more detailed discussion of MPEG-7 see [Kosch2003].

In MPEG-7 Part 5, the Multimedia Description Scheme (MDS), the D and DS dealing with generic features and multimedia descriptions are defined. Along with vector and time descriptors, textual description tools, controlled vocabularies, etc. it offers DS for describing the conceptual aspects of multimedia content, called the Semantic DS. Objects derived from semantic base objects and an extendable set of predefined relations between these objects can be used for constructing a semantic graph, describing the multimedia content. Note that the base objects provide references to time, places, and agents etc. and thus the semantic graph is powerful enough for describing arbitrary multimedia content.

Based on the MPEG-7 Semantic DS an annotation and retrieval prototype for MPEG-7 based image descriptions has been implemented in previous projects[1] (see [Lux2004] for detailed information on the implementation). The annotation tool, called Caliph (from *Common And LIght-weight Photo annotation*), allows the visual construction of MPEG-7 semantic descriptions. Based on an instance catalogue already created or imported instances of semantic objects can be used for creating new descriptions by dragging the semantic object instances onto a drawing panel and interconnecting them visually with arrows, which represent relations.

The retrieval tool, called Emir, allows keyword based retrieval of image descriptions. In a second step a concept for retrieval of semantic graphs was created and implemented, which will be described in detail in the following chapter. Although the prototype is restricted to image descriptions it can be easily extended to support every kind of media supported by MPEG-7 as the fact that the handled items are images is only used for the visual representation of retrieval results.

[1] The prototype called Caliph & Emir is available along with documentation and screenshots at http://caliph-emir.sf.net

3 Retrieval Mechanism for Semantic Descriptions

This section introduces our retrieval model, which is motivated by providing a fuzzy retrieval mechanism for semantic descriptions and an appropriate ranking scheme, including support for wildcards. As there are already well functioning and well tested tools for text retrieval available one major constraint is that we want to focus on the usage of existing text retrieval tools. All of the used techniques should find their source in existing text retrieval techniques to allow the usage of existing tools if possible to rely on their speed and precision.

For the retrieval process of MPEG-7 based semantic descriptions we can assume that, without loss of generality, a set of images exists, where each image is annotated with a semantic description[2].Thus, our goal is to retrieve a set of semantic graphs best matching a given input graph. In the following section nodes (or vertices) of the graph are denoted as semantic objects and edges of the graph are denoted as semantic relations. Our model is described in three parts, whereas the first part explains the indexing of semantic objects and semantic descriptions, the second part states on the retrieval process and the third part introduces the ranking method.

3.1 Indexing of Semantic Descriptions

The first step is creating a data structure for accessing nodes *N* of the graph. As in text retrieval we are using an inverted index as data structure, which holds all semantic objects and offers good possibilities for speeding up the retrieval process as a whole. Although the process cannot be explicitly stated as indexing process it includes an indexing process and has a similar intended purpose.

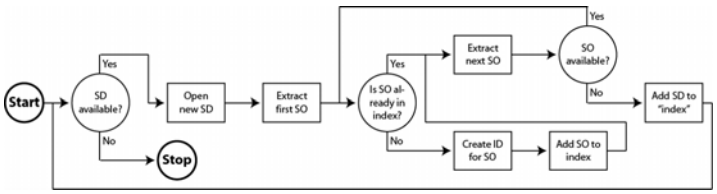

Figure 1 - Flow diagram showing the process of "indexing" semantic descriptions (SD) and semantic objects (SO).

In general for every unique semantic object in all semantic descriptions a unique identifier is assigned and an index entry is created, which means the text describing a semantic object is indexed using text retrieval methods. Note that, multiple semantic descriptions can share the same semantic objects. In this case each shared semantic object is treated as one object and obtains the same unique ID within the different semantic descriptions. Figure 1 shows in detail the implementation of the indexing process,

[2]The semantic descriptions is given in MPEG-7 or in a format easily transformable to MPEG-7.

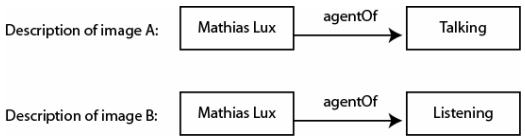

Figure 2 - Examples for semantic descriptions of two different images

For example if the descriptions shown in Figure 2 are processed three different semantic objects with three different IDs are extracted: (i) Mathias Lux (Semantic agent, part of description of image A and B, ID: 1), (ii) Talking (Semantic event, part of description of image A, ID: 2) and (iii) Listening (Semantic event, part of description of image B, ID: 3)

After indexing and assigning unique IDs to semantic objects, semantic descriptions are indexed using a string representation. Given the semantic descriptions in Figure 2, the string representation of image A is: [1] [2] [agentOf 1 2], for image B the description is [1] [3] [agentOf 1 3]. In the first part of the string representation all available semantic objects (vertices) of the semantic description (graph), represented by their ID in square brackets, are defined in numerically ascending order. The second part consists of all available semantic relations (edges) in lexicographically ascending order. Each semantic relation is defined in square brackets, whereas the name of the relation is followed by the ID of its source and the ID of its target. Note that the number of possible MPEG-7 based semantic relation types is already limited by the standard itself. Therefore relations do not get unique IDs but are referenced by their names. All possible semantic relations in MPEG-7 are directed edges but have an inverse relation. Based on this fact relations are re-inverted if their inverse relation is used in a graph. It can be seen that the string representation is unique, but due to the lack of space a formal prove is not given here.

3.2 Retrieval of Semantic Descriptions

Given the fact that above described indices and representations exist, a retrieval mechanism can be implemented as follows:

1. A user provides a query string for each of the *k* semantic objects he wants to search for and interconnects the nodes with relations. Each query string leads to a node query, q_1 to q_k, which are used to query the inverted index of semantic objects, which is described in 3.1.
2. The retrieval engine searches for a set of available matching node IDs L_{q1} to L_{qk} for each node query q_1 to q_k, sorted by relevance of the matches. The relevance returned for each relevant node is in *(0, 1]*, whereas a relevance of *1* indicates an optimal match. The relevance is obtained from using standard text relevance methods (e.g. vector space model).

3. Based on the sets of matching nodes for each node query the original query is expanded to $|L_{q1}|\bullet|L_{q2}|\bullet \ldots \bullet|L_{qk}|$ queries, for which the node IDs and the relevance of the nodes are available. This means that every node returned by q_i is combined every node returned by q_j having $i \neq j$. Given the semantic relations of the user queries consisting of semantic descriptions can be created.
4. For each of the above $|L_{q1}|\bullet|L_{q2}|\bullet \ldots \bullet|L_{qk}|$ queries the number of matching documents is found through a search in the string representations of the graphs with regular expressions. A relevance value for each matching document is calculated based on the formula presented in 3.3.
5. All resulting sets from step 4 are merged in one result set, whereas for documents which are in more than one set, a higher relevance value is assigned.

3.3 Relevance Calculation

Taking one specific expanded query q with node set $N^q = \{n_1^q, n_2^q, ..., n_k^q\} \neq \varnothing$ and relation set $R^q = \{r_1^q, r_2^q, ...r_l^q\}$ and one specific matching semantic description d resulting from the search in 3.2.4 with node set $N^d = \{n_1^d, n_2^d, ..., n_r^d\}$ and relation set $R^d = \{r_1^d, r_2^d, ...r_s^d\}$ with $k, l, r, s \in \mathrm{N} \cup \{0\}$. The relevance $r \in (0,1]$ based on the query nodes relevance values $r(n_1^q), r(n_2^q), ..., r(n_k^q) \in (0,1]$ is defined by

$$r = \frac{\min(|N^q| + |R^q|, |N^d| + |R^d|)}{\max(|N^q| + |R^q|, |N^d| + |R^d|)} \cdot \prod_{i=1}^{|N^q|} r(n_i^q)$$

Equation 1 - Relevance of matching semantic description for one expanded query

The calculated relevance takes the relevance of nodes, which result from the query expansion, into account. The relevance value is in the interval (0, 1] because all node relevance values are in (0, 1] and the fraction has to be in (0, 1] because the numerator is smaller or of equal size compared to the denominator. Note that all irrelevant nodes are discarded in step 3.2.2 by discarding all nodes with relevance below a specific threshold which leads to a minimum relevance above zero. The relevance of semantic relations is not taken into account as the relations in the query are only matched with relations in the database following the Boolean model, not supporting a partial or fuzzy match.

To express the meaning of the relevance formula in words: The more relevant the nodes of the query expansion are, the more relevant is the matching semantic description. Additionally the smaller the difference in the number of components (nodes and edges) of the query and description graph is, the more relevant is the matching semantic description.

3.4 Implementation Details

The above described method was implemented within a pre existing framework called *Emir*, which stands for *Experimental Metadata based Image Retrieval*. Emir uses the Jakarta Lucene search engine [Lucene], which allows the creation of an inverted index of nodes. The string representations of semantic descriptions are stored in a flat file.

For query formulation a simple query language is used which can be described as follows: All nodes are defined in square brackets, inside these square brackets all features of Lucene like fuzzy matching or field matching can be used. Following these node queries the relations are defined using the name of the relation as defined in the MPEG-7 standard followed by the source of the relation and the target of the relation identified by the position in the list of node queries. Following BNF expression defines the supported queries:

```
Query ::= NodeQuery {NodeQuery} {RelationQuery}
NodeQuery ::= “[“ NodeQueryString “] “
NodeQueryString::= ( Clause )*
Clause ::= [“+”, “-“] [<Term> ":"] ( <Term> | "(" NodeQueryString ")" )
RelationQuery ::= <MPEG-7_Relation> <Number> <Number>
```

From each of the expanded queries a regular expression for searching in the file of semantic descriptions is created and executed on each semantic description in its string representation. If the regular expression matches the string, the associated documents are put into the result set and the relevance of the documents is calculated. Finally the result sets are merged following above described parameters and the sorted set of results is presented to the user.

4 Conclusion

A more flexible presentation mechanism with incremental updates of the result list while expanding in the background would be more convenient. In comparison to a naïve approach of indexing all textual content of nodes and querying the index complex concepts like “person A talks to person B” can be retrieved. Another already tested approach would be to present a catalogue of existing semantic objects to the user. This would result in a more complex retrieval process for the user. In a first step the user has to identify the needed semantic objects by browsing the catalogue list or hierarchy to construct the query graph in a second step. In our approach the browsing step is omitted, which speeds up the process and allows the user to be purposeful unspecific, e.g. *I-Know* for a query term matches all I-Know conferences from I-Know ’01 to I-Know ’04.

5 Future Developments

The number of resulting queries after query expansion can become a problem if the query string for a node is too unspecific and returns too many node results. In the current implementation for the query expansion all nodes with a maximum relevance, which is considered to be a relevance value of 1, are taken into account as well as the next matching node. Although this heuristic has shown well behaviour query expansion should be either configurable for the user or should be done with a dynamic threshold. For such a threshold no heuristics have been found yet, different approaches are possible like observing difference between adjacent values in the result list for identifying big drops in relevance.

The presented approach will be evaluated in a next step in comparison to naïve approaches like Boolean matching of semantic graphs and indexing all textual contents within an inverted index. This comparison will be based on a test data set yet to be defined. The retrieval user interface will be updated to support incrementally changing result lists with background query expansion. A further addition is the visualization of semantic graphs and semantic information in the result list by drawing the semantic graphs or presenting a textual summary of the semantics generated by the system.

Acknowledgements

The Know-Center is a Competence Center funded within the Austrian Competence Center program K plus under the auspices of the Austrian Ministry of Transport, Innovation and Technology (www.kplus.at).

References

[Bormans2002] Bormans, Jan, Hill, Keith, "MPEG-21 Overview", Moving Picture Expert Group MPEG, Shanghai, October 2002, URL: http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm

[DelBimbo1999] Del Bimbo, Alberto, "Visual Information Retrieval", Morgan Kaufmann Publishers, 1999

[Flickner1995] Flickner, Myron, Sawhney, Harpreet, Niblack, Wayne, Ashley, Jonathan, Huang, Qian, Dom, Byron, Gorkani, Monika, Hafner, Jim, Lee, Denis, Petkovic, Dragutin, Steele, David, Yanker, Peter, "Query by image and video content: The QBIC system", IEEE Computer, 28(9), pp. 23-32, September 1995

[IBM2004a] IBM Research, "MARVEL: MPEG-7 Multimedia Search Engine", http://www.research.ibm.com/marvel/, last visited: 25.11.2004

[IBM2004b] IBM Research, "VideoAnnEx Annotation Tool", http://www.research.ibm.com/VideoAnnEx/, last visited: 25.11.2004

[Kosch03] Kosch, Harald, "Distributed Multimedia Database Technologies supported by MPEG-7 and MPEG-21", CRC Press, November 2003

[Lossau2004] Lossau, Norbert, "Search Engine Technology and Digital Libraries - Libraries Need to Discover the Academic Internet", D-Lib Magazine, Vol. 10, Num. 6, June 2004

[Lucene] The Apache Software Foundation, "Jakarta Lucene", a Java based search engine, URL: http://jakarta.apache.org/lucene

[Lux2004] Lux, Mathias, Klieber, Werner, Granitzer, Michael, "Caliph & Emir: Semantics in Multimedia Retrieval and Annotation", 19th International CODATA Conference, Berlin, Germany, November 2004

[Martínez2003] Martínez, José M., "MPEG-7 Overview", Moving Picture Expert Group MPEG, Pattaya, March 2003, URL: http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[Mayer2004] Mayer, Harald, Bailer, Werner, Neuschmied, Helmut, Haas, Werner, Lux, Mathias, Klieber, Werner, "Content-based video retrieval and summarization using MPEG-7", in Proceedings of Internet Imaging V, IS&T/SPIE 16th Annual Symposium, Electronic Imaging, San Jose, California USA, 2004

[Smeulders2000] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R. "Content-based image retrieval at the end of the early years", Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 22, No. 12, pp. 1349-1380, December 2000

[Summann2004] Summann, Friedrich, Lossau, Norbert, "Search Engine Technology and Digital Libraries - Moving from Theory to Practice", D-Lib Magazine, Vol. 10, Num. 6, September 2004

[Tsinaraki2003] Tsinaraki, Chrisa, Fatourou, Eleni, Christodoulakis, Stavros, "An Ontology-Driven Framework for the Management of Semantic Metadata Describing Audiovisual Information", in Proceedings 15th Conference on Advanced Information Systems Engineering CAiSE 2003, pp. 340-356, Springer, LNCS, 2003

[Wactlar2002] Wactlar, Howard D., "Extracting and Visualizing Knowledge from Film and Video", in Proceedings 2nd International Conference on Knowledge Management I-KNOW ’02, Journal of Universal Computer Science, July 2002