Semi-Automatic Annotation and Retrieval of Visual Content Using the Topic Map Technology

Atta Badii, Chattun Lallah, Oleksandr Kolomiyets, Meng Zhu, Michael Crouch
IMSS, Intelligent Media Systems and Services Research Laboratory
University of Reading
UNITED KINGDOM

Abstract: - There are still major challenges in the area of automatic indexing and retrieval of multimedia content data for very large multimedia content corpora. Current indexing and retrieval applications still use keywords to index multimedia content and those keywords usually do not provide any knowledge about the semantic content of the data. With the increasing amount of multimedia content, it is inefficient to continue with this approach. In this paper, we describe the project DREAM, which addresses such challenges by proposing a new framework for semi-automatic annotation and retrieval of multimedia based on the semantic content. The framework uses the Topic Map Technology, as a tool to model the knowledge automatically extracted from the multimedia content using an Automatic Labelling Engine. We describe how we acquire knowledge from the content and represent this knowledge using the support of NLP to automatically generate Topic Maps. The framework is described in the context of film post-production.

Key-Words: - Topic Map Technology, Natural Language Processing, Automatic Labelling, Knowledge Representation

1 Introduction
The DREAM (Dynamic Retrieval Analysis and semantic metadata Management) project aims at paving the way towards semi-automatic acquisition of knowledge from visual content which nowadays is growing at an increasing rate. This is a DTI Funded project, being undertaken in collaboration with Partners from the UK Film Industry, including Double Negative [3], The Foundry [8] and FilmLight [4]. Double Negative is the test partner who provided the test materials and user requirements and evaluations. One of the main issues in this industry is the storage and management of huge repositories of multimedia data, in particular, video files, and being able to search through scattered repositories when there is a need for a particular video shot. For example, when Special Effects Designers need a category of clips containing “fire explosions” which they can use in the making of a new special effect, it is a tedious and time consuming task for them to search for similar video clips which have some specific objects of interest. The first prototype of DREAM has been evaluated in this post-film production application domain and aims at resolving the existing indexing and retrieval problems of video clips.

In current indexing and retrieval systems, the associated knowledge in video clips is mainly characterised by keywords which usually do not depict the semantic content of those clips. In some cases the knowledge associated with the objects of interest in a video clip are captured, but the enhanced knowledge using the contextual information in relation to other similar objects is not available. This cannot possibly be scaled up to meet the increasing demands of current and future information accessibility to serve personal and professional needs, in particular film post-production which requires support for a gigantic amount of multimedia data. There are also many other unresolved research challenges for a video database and query support system. These range from video meta-feature/genre-specific classification issues to subject and cinematographic properties level of indexing, query languages design, man-machine cooperative data analysis and labelling platform as well as performance measurement, and including efficient implementation of functionalities that support storage, recognition and retrieval of the video files.

We believe that the most promising route to innovation in this area that can meet the above challenges lies in advanced ontology network engineering underpinning a semantic architecture that would provide semantic and context-aware bridges across the various objects and features of a video clip. This will enable users to search through the semantic content of those video clips. The DREAM Project also aims to break new grounds by introducing and implementing the concept of evolving
multimedia ontologies for visual data using new concepts like Topic Maps [1]. It links the concept extraction from multimedia with ontology evolution, creating a synergy of enormous yet unrealized potential. Our proposed framework will address the needs of individual goals such as retrieval for leisure or to serve routine business processes in film post-production, security and forensics video stream scrutiny or medical (tele-) healthcare applications.

In this paper, we discuss the general framework of DREAM, describing the work flow between the different DREAM modules. We then described the Topic Map Technology and explained how automatic Topic Map population is carried out in this project. The paper emphasises how we extended the use of Natural Language Processing for Knowledge Acquisition and how Topic Maps are used as a tool for Knowledge Representation.

2 The DREAM Framework

The main challenge in this project is to architect an indexing, retrieval and query support framework which can exploit content, context and search-purpose knowledge as well as any other domain related knowledge in order to ensure robust and efficient multimedia object labelling, indexing and retrieval using Topic Map Technology. To meet this challenge, our current Framework enables semantically triggered human intervention to support optimal cooperation in the semi-automatic labelling of what is currently mostly a manual process. This framework is underpinned by a network of scalable ontologies, which grows with the continuous ongoing annotation of video content. To support these scalable ontologies, we deployed the Topic Map Technology, which also enables transparent and flexible multi-perspective access to the knowledge.

Figure 1 shows the main modules of the DREAM Framework modelled for the film post-production domain. In this domain, we have huge repositories of video clips, which may reside in different locations. Using the above framework, the DREAM User will be able to query or navigate through the repository using the Topic Map Visualisation Engine which provides a single interface to query the knowledge base or to navigate through the connection of the semantic concepts of the video clips. To support this query and navigate through the knowledge base, a Java Topic Map Engine has been designed and implemented. This uses the power of Java Hibernation for the dynamic storage, updates and querying of semantic visual metadata.

To construct those metadata, a Collaterally-Cued Automatic Labelling Module has been implemented. This reads in the video clips and extracts the main objects of interest, in terms of semantic keywords, from the clips. The DREAM User can confirm those keywords and/or add more contextualised or domain-specific information to then make a set of keywords, which is then fed into the NLP Engine. This Engine uses external knowledge such as WordNet to add meaning to the captured information. This enhances the knowledge, which are then in a form we term as Semantic Containers, these are passed to the Topic Map Engine, which merges them with the existing knowledge found in the DREAM Topic Map Database. This enables intelligent querying and visualisation of the Semantic Network of concepts which indexes millions of video shots.

3 Topic Map Technology

Topic Maps, as defined in ISO/IEC 13250 [5], is an international standard for organising and representing information on the Web. A topic map is basically an XML document in which different element types are used to represent topics, occurrences of topics, and associations (or relationships) between topics. The Topic Maps model provides a mechanism for representing large quantities of structured and unstructured information and organising it into “topics”.

---

Fig.1. DREAM Framework for film post-production domain
The topics can be related by associations and can have occurrences linked to them. A Topic Map can thus be referred to as a collection of topics and associations. The associations are used to navigate the information in many different ways. The network can be extended to grow with the collection, or can be merged with other topic maps to provide additional paths through the information. The most important feature of Topic Map is that it explicitly captures additional tacit information. It is this capability that has captured the interest of people working on knowledge management issues. They see Topic Maps as a mechanism that can help capture what could be considered “knowledge” from a set of information.

Although the creation of Topic Maps is quite simple, using the basic concepts – topics, their occurrence and associations, as a model of knowledge representation, the Topic Maps model is based on three main issues:
1. extraction of topics (subjects) from structured and unstructured data,
2. defining associations (relationships) among topics and
3. linking those topics with a data layer (resources)

Automatic Topic Map creation/population is still an unresolved problem among the Topic Map Community. Manually creating a Topic Map is a tedious job, time consuming and expensive. The automatic generation and population of Topic Maps have two main driving points. These are, firstly, the need to integrate existing information resources into a new Topic Map structure, and secondly the desire for on-the-fly Topic Map creation, from input information, without constant human intervention. In the first case, an automatic solution is certainly desirable, as the time cost of having to create a vast number of Topic Maps would be very expensive. The key to such automatic generation of Topic Maps is the input data. If the input data is structured in a manner which could be easily translated to a Topic Map structure, then this process is greatly simplified. However, when the data resource is something abstract or unstructured, such as free text (with no mark-up), then the process of automatically representing this data in a Topic Map format, becomes more complex.

One example of automatic Topic Map population from abstract data, such as free text, was presented at TMRA2007 (Topic Maps Research and Applications International Conference). This project presents the use of linguistic templates for automatic Topic Map generation from free text [2]. It uses a set of linguistic templates on the input text, to extract topics, associations and occurrences, automatically. These automatically generated Topic Maps were a good base for human editing to build upon. However, until now, no break through solution has been proposed which claims to automatically generate a knowledge base, with the least amount of human intervention.

Having a solution to automatically generate a Topic Map from raw and unstructured data is required for the adoption of the Topic Map Technology by a wider user community. In the DREAM project, we extend the use of Natural Language Processing to improve video clip indexing and provide a powerful tool for concept extraction from free text. The automatic Topic Map generation in DREAM is quite innovative, as described in Section 4.

4 Knowledge Acquisition

Knowledge acquisition is carried out using the DREAM NLP Engine. The automatic Topic Map generation in DREAM uses Natural Language Processing (NLP) techniques to process free text (and/or keywords). Those keywords are extracted by the Automatic Labelling Engine and confirmed by the user. The Automatic labelling Engine aims at automatically assigning semantic keywords to objects of interest appearing in the video segment. The module had been implemented to label the raw video data in a fully automatic manner. The user of the system will be able to use the module to facilitate the labelling and indexing of the video data. With this function, all the objects of interest including moving and still foreground objects will be labelled with linguistic keywords, which the user can confirm or revise.

![Fig. 2, Workflow of the automatic labelling engine](image)

The Automatic labelling Engine takes the low-level blob-based visual features, i.e. colour, texture, shape, edge, motion activity, motion trajectory, as input to compare with the visual concepts defined in the visual vocabulary of objects, which consists of a set of clusters of visual feature vectors of different types of special
effects foreground objects like blood, fire, explosion, smoke, water splashes, rain, clouds etc, to find best matching visual concepts using K-Nearest Neighbour algorithm. 453 film post-production special effect video clips were selected from the industrial library dataset as our training and testing corpus. A visual vocabulary consisting of 32 visual concepts had been constructed by clustering the feature vectors extracted from the training dataset. The Automatic Labelling Engine can successfully identify the testing images as one of the 32 semantic categories, e.g. fire, explosion, water drip, blood, gore, etc. The average labelling accuracy of the tagging result is above 90%.

The NLP Engine makes use of the OpenNLP package [7] and implements NLP templates for extraction of concepts to populate a Topic Map. Besides the mandatory syntactic analysis such as sentence detection, tokenisation and parsing, the engine carries out shallow semantic analysis of the natural language input. As the DREAM project targets a specific application domain, we followed a heuristic-based approach, in which we carried out a manual analysis of video annotations and developed a set of customised templates for recognition of objects of interest as well as their thematic roles. In doing so, we not only recognised concepts for automated population but also utilized the linguistic information to estimate the importance of the concepts.

A natural language statement can be seen as a description of an event in terms of what happened to whom, when and where. Early information extraction approaches targeted at filling slots in a predefined pattern by recognising phrases corresponding to these semantic categories. The pattern, in turn, follows the semantics enclosed in the corresponding verb. Later, the pattern can be used for information retrieval and data access purposes. By contrast, NLP support in DREAM is much more flexible and does not rely on predefined patterns, but rather creates semantically sound constructions by providing links between sentence constituents on the fly, which all, in turn, point to a video sequence being annotated. Despite the many ambiguities in language, such an approach has a number of advantages in comparison to simple keyword-based annotations. Since the application domain DREAM is designed for is quite narrow with a small vocabulary, the keyword-based approach for retrieval would not have good discriminative coverage and will lead to overloaded result sets with low precision. By contrast, a natural language interface for information retrieval allows the user to query a data source by phrases and statements that are not restricted to a word vocabulary.

Technically, the DREAM NLP engine makes use of the results of syntactic analysis done by OpenNLP, and creates a flat structure representing an event and comprising a number of recognised concepts or even a number of nested events. It also makes complete use of the complementary linguistic information such as adjectives and adverbs providing the scope for the concepts they modify. A single event is represented as a structure involving two topics, which represent concepts extracted from text with a governing verb, which denotes the event, for example, “John is reading a book”, is represented as a frame reading[“John”, “a book”].

Other linguistic phenomena such as adjectives and adjectival phrases are also captured and since they are syntactically governed by a high-level noun phrase, these are saved as modifiers of the head noun. Later, the modifiers are added to the Topic Map as scope for the topic. For example, the noun phrase “a big red car” is captured as the topic “car” with the scope [“big”, “red”]. Prepositional phrases in textual video annotations were first manually analysed with a conclusion that in video annotations they are attached to a verb in a sentence. Thus, all prepositional phrases are transformed in a way that multiplies the verb with all the prepositions in the sentence composing relationships such as “landing at”, “landing on” which semantically denote the different association roles of their members.

We also consider more complex sentences, which do not have direct objects, but involve another sub-sentence. In order to be able to handle such a phenomenon we introduce a new Semantic layer to Topic Maps, which can represent an entire sentence, which in turn can be used as a topic for another relationship in the Topic Map. An example of a complex sentence is this sample text from the BBC website: “No10 denied that Gordon Brown exploited Lady Thatcher for political benefits”. Processing this sample sentence using current NLP techniques to extract topics and associations will be almost impossible. The most robust NLP algorithms will most probably identify the following topics (entities) and associations (actions):

- Topics: “No10”, “Gordon Brown”, “Lady Thatcher”, “Political benefits”
- Associations: “Deny”, “Exploit”

However, the action “Deny” is associated to a cloud of topics: [No10] [denied] [Gordon Brown exploited Lady

\[\text{Nested prepositional phrases starting with “of” are not transformed in this way and together with the head nouns form a topic name, e.g. “the city of London”}\]
Thatcher for political benefits]. “No10” is denying an event and it is very difficult to extract this semantic construction and represent this in a Topic Map. To be able to handle this complexity, we introduced a Semantic Layer in the Topic Map Generation process, as shown in Figure 3. This layer allows a topic to be related to an event (a group of related topics).

The core outputs of the DREAM NLP Engine are termed “Semantic Containers”. These semantic containers allow the representation of both simple sentences and complex sentences in terms of entities and actions, which are used by the DREAM Topic Map Engine to generate Topic Maps automatically. A semantic container may contain other semantic containers as shown in Figure 4. This enables an entity (topic) to be linked to a group of entities (topics). As a result, complex sentences are represented by a single semantic container, with a number of “inner” semantic containers detailing the semantic information processed by the NLP Engine.

5 Knowledge Representation

The DREAM Knowledge Base exploits the use of the Topic Map Technology to create a rich knowledge base, which consists of Upper Ontologies, Domain Specific and Domain Generic Ontologies, as illustrated in Figure 5. The Semantic Containers as described in section 4 are used to populate the DREAM Knowledge Base. Entities and Actions extracted from those Semantic Containers are used to create the topics and associations, which are located above the mid-level ontologies and upper ontologies. The new topics which are added to the Knowledge Base are automatically merged with this ontology. This makes the Knowledge Base dynamic and allows it to continually grow with the addition of new topics extracted from new video shots. The Knowledge Base is in other words, enriching itself and is eventually thereby enabling a more intelligent search. Such as for example, if a user searches for “blood”, the system will know that “blood is a kind of “gore”, and will also return the different types of “gore”. This provides more intelligent results to the user. Also the DREAM Visualisation Engine allows uncomplicated navigation through the connection of topics and enables the user to easily find the video clip he/she is looking for, as all the topics in the Knowledge base have a set of occurrences, which index the video clips.
Topic Map allows casual browsing of the knowledge base with a richly cross-linked structure over the repository content. Topic occurrences create ‘sibling’ relationships between repository objects, the video shots. A single resource may be the occurrence of one or more topics, each of which may have many other occurrences. When a user finds/browses to a given resource, this sibling relationship enables them to rapidly determine where the other related resources are to be found. Topic associations create ‘lateral’ relationships between subjects, the movie concepts – allowing a user to see which other concepts covered by the repository are related to the subject of current interest and to easily browse these concepts. Associative browsing allows an interested data consumer to wander across a repository in a directed manner. A user entering the repository via a query might also find associative browsing useful in increasing the chance of unforeseen discovery of relevant information. A DREAM Topic Map Visualisation, as shown in the Framework diagram [Fig.1] has been implemented to provide such interactive visual browsing of the DREAM Knowledge Base.

6 Conclusion
In this paper, we have presented the DREAM Framework and discussed how it addresses some of the problems of efficient indexing and retrieval of video content. The framework architecture, which has been presented in the context of film post-production, responds to the challenges of efficient semantic-cooperative retrieval. We described how this framework exploits the use of NLP to perform automatic Topic Map population and creating a self-evolving semantic network for any media repositories, by defining the topics (concepts), especially in video. We also discussed how collaborative labelling is handled through the Automatic Labelling Engine. The first DREAM prototype has already been implemented and allows the browsing (and editing) of Topic Maps, through topics and associations, to locate video shots.

References: