Towards a Semantic Oriented Multimedia Indexing and Retrieval Solution

Mihaela Brut

Abstract—**The main purpose of the present article is to discuss a solution for the efficient semantic-oriented multimedia information retrieval with the support of both the traditional multimedia indexing and classification techniques (based on processing the multimedia content itself) and the modern semantic Web technologies (oriented to exploring the associated multimedia metadata). The challenge proposed by this research concerns a solution for the problem of bridging the "semantic gap" between low-level multimedia features and high-level concepts describing the content. The solution propose to organize the various metadata types situated in this interval into a three layers structured collection: generic metadata, valid for all multimedia types; media-specific metadata; semantic metadata for describing the multimedia objects content with the support of various specific vocabularies and ontologies. Supplementary semantic information should be managed through a knowledge database. In order to integrate the existing heterogeneous set of automatic indexation and classification algorithms, a structured algorithms collection should be defined and developed. A generic interface for the algorithms should be adopted, where their input/output data, as well as their preconditions and effects will be correlated with the structured metadata collection. Alongside to searching into metadata collection, three issues should be considered when providing results to the user complex queries: to process the query, to define the algorithms sequences in order to be successively applied for indexing the multimedia content, and to use the information provided by the knowledge database. For gaining a user-centered dimension, our approach considers a user profile expressing his long-term interests and his current activity goals in terms of multimedia structured metadata. The profile is useful for refining the user query's results and to provide him with personalized recommendations.**

Index Terms—**multimedia indexing, semantic metadata, information retrieval.**

I. Introduction

THE main purpose of the present article is to discuss the issues raised for developing a solution for the efficient semantic-oriented retrieval of the multimedia information belonging to multidisciplinary areas, based on the traditional multimedia indexing and classification techniques as well as on the modern semantic Web technologies. The main focused issues are:

- how to develop an solution for organizing the large palette of the multimedia metadata types, from the low-level multimedia features to the semantic metadata based on ontology constructs;
- how to develop a flexible solution in order to obtain a structured collection by organizing the existing set of automatic indexation and classification algorithms for textual documents, images, audio and video content. This structure should enable to convert the algorithms results to the defined metadata structure and to compose certain algorithm sets.
- how to develop a multimedia retrieval solution as response to the user's semantic queries by processing the queries, applying multiple algorithms, using a supplementary knowledge database;
- how to develop a user profile according the user preferences and activities; this profile should be used to improve the query results, according user characteristics.

The challenge proposed by this article concerns a solution for the problem of bridging the "semantic gap" between low-level multimedia features and high-level concepts describing the multimedia content. Since low-level features do not encapsulate the high-level semantics of a document, the development of systems which automatically extract rich semantic descriptions involves a interdisciplinary approach, combining complex techniques such as content analysis, knowledge databases, machine learning, and semantic Web (more recently), and a general solution is far from being achieved.

Various domains such as news gathering, TV, banks of resources for commercial or consumer applications, collaborative work, video surveillance were flooded in the last years by a huge amount of video and multimedia sources; now, these domains are in a growing demand of solutions for their management. The most sensitive aspect concerns the multimedia content semantics, and big efforts were accomplished for acquiring its transparency for the computer applications.

Semantic Web technologies were adopted at a large scale by the social Web applications, where the main solution for semantically indexing huge multimedia collections is based on the free *tagging* activity performed by the users

M. Brut is Lecturer PhD at the Alexandru Ioan Cuza University, Faculty of Computer Science, Iasi, 16 Berthelot Street, 700483, Romania (phone: +4-0232-201544, e-mail: mihaela@info.uaic.ro).

themselves[1], which raises the difficulty of keeping the annotations consistency. General semantic Web applications adopt retrieval mechanisms based on processing not the effective resources content, but their associated metadata: free tags developed through the community efforts, microformats, metadata in various specialized formats (RDFa, DCMI: Dublin Core Metatada Initiative, FOAF: Friend Of A Friend, RSS: Rich Site Summary etc.), or even ontology-based metadata (expressed in RDF: Resource Description Framework or OWL: Web Ontology Language) [5].

The traditional multimedia indexation and classification techniques are, by contrary, focused on the effective multimedia content processing. The indexation algorithms for images, audio and video content are mainly in charge with low-level multimedia features analysis, while the text indexation algorithms and the general multimedia classification algorithms concern also some semantic features regarding the content description, but not in terms of high-level concepts (such as ontology or vocabulary concepts).

The present article is focused on possible directions for developing a solution for retrieving the multimedia content belonging to multidisciplinary areas, by combining technologies from both semantic Web and multimedia indexation/classification domains.

In the next sections, we will present the current state-of-the art and possible solutions for organizing the multimedia metadata, as well as the existing multimedia indexation algorithms. We will then present the main existing approaches in the multimedia retrieval area and our proposed approach in the context of the previously exposed metadata and algorithms organization solutions. Further, a small discussion on how to integrate user personalization in the retrieval mechanism will be presented. In final, the conclusions and further work directions will be exposed.

II. Multimedia Metadata Organization

From the representational point of view, the main goal of transforming multimedia materials into machine-compatible content is ensured both by the Semantic Web activity of the Web Consortium[2], and by the ISO's efforts in the direction of complex media content modeling, in particular the Multimedia Content Description Interface (MPEG-7)[3]. The two directions are syntactically and semantically different [14], and some solutions to unify them were proposed: a semantic Web approach which uses the schemata developed in MPEG-7 as third-party specifications, or multiple MPEG-7 translations into RDF and OWL. Moreover, a set of tools were developed to enable the automatic extraction of the visual features in multimedia materials (as MPEG proposes) and the manual association of them with ontology concepts (as a semantic web approach requires): PhotoStuff, AKTive Media, Vannotea, M-OntoMat-Annotizer, SWAD etc.[4]. Efforts were made also towards a multimedia annotation interoperability framework[5] and towards a common multimedia ontology framework, aiming a uniform use of the multimedia ontologies, according to their intended use and context [6].

Instead of converting and expressing all metadata types through ontology constructs, our solution is focused on selecting the most suitable metadata types for each content descriptor, and for intelligently organizing them into a structured collection, with three levels: generic metadata, valid for all multimedia types (such as file name, title, author, creation date, last modification date, size); media-specific metadata; semantic metadata expressed through various domain specific vocabularies and ontologies. This third level includes support for addressing the above mentioned “semantic gap”, e.g. for supplementing the meaning produced by the first two levels. For example, <dc:title> generic metadata from DCMI vocabulary only indicates the content as being the title of the current resource, but nothing about the meaning of this title. The solution consists in correlating this low-level metadata with certain ontology constructs [11], as the semantic metadata level proposes, or in considering a knowledge database [7].

III. Managing the Multimedia Indexation Algorithms

In order to produce concrete metadata organized according to the above mentioned structure, automatic indexation and processing activities should be ideally considered. We provide below a short overview of the algorithms and techniques specific to different multimedia document types.

For *textual documents*, some indexing techniques are inspired by the classic Information Retrieval (IR) [16], or by *Web Information Retrieval* (Web-IR), exploiting the hypertext features, such as page hyperlinks [2] and HTML general tags [1]. The progress from a term-based to a concept-based document indexation was possible due to the latent semantic indexing technique [17] or to some models and methods for knowledge representation typical to the artificial intelligence field, such as neural networks, semantic networks, bayesian networks [13].

An *image content semantic indexation* process analyses object related information (e.g. how many objects are in this image?, is object X present?, which objects are in the image?), using an a priori knowledge on the observed scenes and a model of the world. To achieve this, current methods

[1] Such examples are www.youtube.com, www.fickr.com, www.twine.com/, www.slideshare.com

[2] http://www.w3c.org/sw/

[3] ISO/IEC JTC1/SC29/WG11 N4031. MPEG-7 (2001); MPEG-7 XM: http://www.lis.ei.tum.de/research/bv/topics /mmdb/e_mpeg7.html; ISO/IEC JTC1/SC29/WG11 N5527, MPEG-7 Profiles, 2003, Pattaya, Thailand

[4] http://www.w3.org/2005/Incubator/mmsem/wiki/Tools_and_Resources

[5] http://www.w3.org/2005/Incubator/mmsem /XGR-interoperability/

are based on the following steps: feature extraction, clustering/segmentation, object descriptor extraction, object recognition [4]. Pattern recognition techniques, such as boosting or cascade of classifiers execution, have been also applied for image semantic indexation.

Audio analysis is accomplished in some main directions [8]: segmentation (splitting an audio signal into intervals, minding the semantics or the sound source composition), classification (categorizing audio segments according to predefined semantic classes, such as speech, music, silences, background noise, and further sub-classes) and retrieval by content (using similarity measures, to retrieve audio items that perceptually or semantically match the query).

In the area of *Content-Based Video Indexing and Retrieval* [7], many research efforts are focused on automatic techniques for extracting metadata to describe the content of large video data archives. Some of these notions have been adopted in video standards such as MPEG-4 (main profile) and MPEG-7. Because the variety of video content is so large, the quality of the meta-data is quite broad. Content-based video archiving and retrieval systems typically use (and automatically detect) metadata elements such as content type, shot-boundaries, audio keywords, textual keywords, and close-caption text keywords. Activities within video scenes are rarely considered. There is some superficial work on video indexing based on matching scene colour profiles, and some on simple spatial-temporal features of the video mostly from the MPEG-4 development community.

Another video indexation paradigm involves describing the content of video scenes in terms of the activities of people or objects within those scenes – then allowing a user to create queries about those activities. At its simplest, this involves recording the trajectory of every object. More sophisticated analysis algorithms can extract additional data such as the colour scheme of each object (if colour information is present), the class of each object (human, vehicle, type of vehicle, animal, etc), and any specific activity-related information about each object (carrying a bag, raising arms in the air, gait – running, walking, etc.) [18]. Pattern recognition techniques are making progress for several years. Multi-parts techniques have improved performances in case of occlusions. Boosting and cascade of classifiers proved very good and rapid results of detection [19].

This large pallet of multimedia indexation algorithms is partially publicly available, but spread out inside many tools, web applications or projects. The Large-Scale Concept Ontology for Multimedia (LSCOM)[6] organizes more than 800 visual concepts for which extraction algorithms are known to exist. Among such processing tools, Lucene provides support to analyze, index and query the textual documents (lucene.apache.org). Octave provides indexation and processing algorithms for texts, images and audio documents (octave.sourceforge.net). SineQua includes shape detection and speech-to-text extractors for video documents (www.sinequa.com/en/solutions.html). LabelMe annotation tool (labelme.csail.mit.edu) manages a large database for research on object recognition, providing a search interface for objects inside images and for scenes inside videos. Joint result of three European projects, WebLab (weblab-project.org) is an open platform for processing documents, including a set of multimedia processing services and an advanced search interface.

Among the tools with support for classification algorithms, the Weka toolbench for machine learning and data mining (www.cs.waikato.ac.nz/~ml/weka) or RapidMinner (www.rapidminer.com) could be mentioned. For the particular case of the support vector machines classification, SVMlight tool (svmlight.joachim.org) is available together with a special version for structured information classification, SVMstruct (svmlight.joachim.org/svm_multiclass.html), which could be very useful for our proposed metadata structure. Some libraries and APIs are also available for processing the matrix representation of documents (Jama: math.nist.gov/javanumerics) or the RDF-based representation of documents (such as Jena API – jena.sourceforge.net).

As the brief state-of-the-art illustrated above shows, the publicly available algorithms for multimedia documents indexation and classification are characterized by a great heterogeneity concerning their input and output data, preconditions and effects, implementation details, hosting platforms or architectures. Our solution is to develop a generic interface for these algorithms and to organize them into a structured and easy to handling collection.

IV. Towards a Multimedia Retrieval Solution

Usually, a user complex query involves more than a single algorithm to be combined and sequentially run over the multimedia content. In the multimedia-indexing domain, the indexation process for responding to such complex queries is produced by a sequence of more or less sophisticated tools in a given order. But multimedia indexing tools are developed and distributed by different teams with really specialised skills on a single media type, making their exploitation for cross media analysis a challenging issue in multimedia indexing. Collaborative multimedia indexing researches have focused on using a static, manually built chain of multimedia tools to generate an expected index, as the Dutch *Acoi* project proposes (monetdb.cwi.nl/acoi).

When the algorithms themselves are available instead of the tools, a set of rules for proper algorithms sequential combination should be established [9]; the previously

[6] LSCOM project: http://www.ee.columbia.edu/dvmm/lscom/; LSCOM ontology: http://www.lscom.org/

mentioned generic interface for the algorithms is essential for defining such rules. Two main approaches are oriented towards defining such a generic interface. The COMM project (comm.semanticweb.org) proposes a Core Ontology for Multimedia which enables to describe the multimedia types and their particular features. In addition, some patterns for indexing operations are considered, including the semantic annotation pattern and the indexation algorithm pattern, which propose such generic interface. The second approach is the Web Consortium's Algorithm representation use case[7], which proposes algorithm ontology to record and uniformly describe available algorithms for image analysis.

In defining the *algorithms generic interface* in our solution, the input and output data should be considered (and expressed in terms of structured metadata), but also algorithm precondition and effect. The input is always a multimedia object, but the output could be a multimedia object as well as a numerical, Boolean or string value. The algorithm precondition concerns some constraints to be fulfilled for enabling the algorithm application. The algorithm effect should be mentioned for example when the input media object suffers an alteration, as in the color image segmentation. Such generic interface is useful both for gaining a uniform description of the algorithms and for defining algorithms combination rules.

The goal of automatic indexing the semantic content of multimedia data is not reachable only through algorithms combination. In addition, the general technique requires developing and using a *rule base* or a *knowledge base* in order to extract suitable features from the raw multimedia data, to match content, to analyze queries, and so forth [4].

For example, in [10], a hierarchy for the radiology domain (concerning the radiological shapes and their semantics) is used in order to improve search efficiency in radiological databases. In [15], the image features specific to certain content are established, and the hierarchical relation between the primitive color regions and the semantic contents is captured as a *state transition model*. Another example of combining video segmentation with semantic indexing is provided by [11]: after region segmentation, features extraction and object identification, *ontology* of objects, events and concepts is used in order to generate the current video's set of objects, events and concepts, together with their related frame list.

The semantic multimedia indexation and retrieval domain constitute an active research and exploration subject for many European projects, but no solution has emerged as a standard to date. *K-Space* (http://kspace.qmul.net:8080/kspace/) is focused on semantic inference for semi-automatic annotation and retrieval of multimedia content, integrating three research clusters: Content-based multimedia analysis, Knowledge extraction, Semantic representation and management of multimedia. The metadata organization and the reasoning mechanism for acquiring multimedia semantics will constitute a use case for our project. *Vitalas* project (vitalas.ercim.org) is focused on intelligent access to multimedia professional archives, developing solutions for cross-media indexing and retrieval, large scale search techniques, visualization and context adapting. How different media-specific metadata are considered in the retrieval process should be investigated in future. The *Candela* project (www.hitech-projects.com/euprojects/candela) was focused on Video Content Analysis in combination with networked delivery and storage technologies. Results could be used in developing rules for video indexing algorithms combination. *Isere* (Inter-media Semantic Extraction and Reasoning) project aimed to study unified merging multimedia analyses in order to enhance the identification of the semantic content of each elementary medium. The *Muscle* network of excellence (www.muscle-noe.org) was focused on multimedia data mining and machine learning technologies, providing a set of showcases for object recognition, content analysis, automatic character indexing, content-based copy detection unusual behavior detection, movie summarization, human detection, speech recognition, etc.

V. ADDING PERSONALIZATION TO THE RETRIEVAL MECHANISM

Alongside with low-level metadata provided by the automatic indexing algorithms, some vocabulary-based semantic metadata are managed by many such tools and projects, but a standardized methodology for integrating semantic Web technologies in the multimedia indexation and retrieval field is far from being achieved. Our research do not intends such standardization, but discussed how to integrate the existing indexation and classification algorithms into a uniform framework where the multimedia semantics is captured in terms of standardized metadata formats and vocabularies. The multimedia retrieval activity involves both metadata collection querying and new multimedia semantic analysis performing.

For gaining an user-centered dimension, as the actual social Web applications promote, the development of a user profile should be considered in order to be used for improving the query results. In the general adaptive hypermedia systems, the user model includes information about some user features (knowledge, interests, goals, background, and individual traits) [3]. In case of adapting a query results to a specific user characteristics, the user goals are inferred from the user searching or navigational activity [12]. If possible, such systems also include in the user profile some preferences illustrating his interests. Our idea is to adopt such two-layer user model (interests and goals), and to express each of these layers into a similar manner with the structured collection of metadata associated to multimedia objects. Thus, the role of

[7] http://www.w3.org/2005/Incubator/mmsem/wiki/Algorithm_representation_Use_case

the metadata collection increases, and a collaborative filtering approach (as also the actual social Web applications promote) is leveraged: similarities between users could be exploited in order to provide them with search results according to the similar users' best rated items. Moreover, the structured metadata representation of both items and users is a top issue in the adaptive hypermedia systems domain itself, and an approach adopting structured support vector machines for multimedia objects classification seems to be an interesting investigation field, with applications in multidisciplinary and interdisciplinary approaches.

VI. Conclusion

The main purpose of the present article was to discuss a solution for the efficient semantic-oriented multimedia information retrieval with the support of both the traditional multimedia indexing and classification techniques (based on processing the multimedia content itself) and the modern semantic Web technologies (oriented to exploring the associated multimedia metadata). The challenge proposed by this research concerns a solution for the problem of bridging the “semantic gap” between low-level multimedia features and high-level concepts describing the content. The solution proposed to organize the various metadata types situated in this interval into a three layers structured collection: generic metadata, valid for all multimedia types; media-specific metadata; semantic metadata for describing the multimedia objects content with the support of various specific vocabularies and ontologies. Supplementary semantic information should be managed through a knowledge database. In order to integrate the existing heterogeneous set of automatic indexation and classification algorithms, a structured algorithms collection should be defined and developed. A generic interface for the algorithms should be adopted, where their input/output data, as well as their preconditions and effects will be correlated with the structured metadata collection. Alongside to searching into metadata collection, three issues should be considered when providing results to the user complex queries: to process the query, to define the algorithms sequences in order to be successively applied for indexing the multimedia content, and to use the information provided by the knowledge database. For gaining a user-centered dimension, our approach considered also a user profile expressing his long-term interests and his current activity goals in terms of multimedia structured metadata. The profile is useful for refining the user query’s results and to provide him with personalized recommendations.

Based on the exposed general solution, we will focus in our future research on the details of each particular issue in order to develop a semantic oriented multimedia retrieval system.

References

[1] Agosti, M., Smeaton, A.F. (eds.): *Information Retrieval and Hypertext*. Kluwer Academic Publishers, Dordrecht, NL, 1997

[2] Brin, S. and Page, L., The anatomy of a large-scale hypertextual web search engine. Comp. Networks and ISDN Systems, 30 (1–7) (1998)

[3] Brusilovsky, P., Millán, E., User Models for Adaptive Hypermedia and Adaptive Educational Syst., in P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), *The Adaptive Web*, LNCS 4321, Springer, 2007

[4] Chen, Shu-Ching, Kashyap, R. L., Ghafoor, Arif, *Semantic Models for Multimedia Database Searching and Browsing*, Kluwer Academic Publishers, NY, 2002

[5] Daconta MC., Smmith, KT., Obrst, LJ, The Semantic Web: A Guide to the Future of XML, Web Services and Knowledge Management, John Wiley & Sons, 2003

[6] Di Bono M.G., Pieri, G., Salvetti, O. (2004), “A Review of Data and Metadata Standards and techniques for representation of Multimedia Content”, MUSCLE project deliverable, IST-CNR

[7] Donderler, M. E. Saykol, E., Arslan, U., Ulusoy, O., Gudukbay, U., BilVideo: Design and Implementation of a Video Database Management System, *Multimedia Tools and App.*, 2008.

[8] Jonathan Foote, “An overview of audio information retrieval”, in *Multimedia Systems*, vol. 7 no. 1, pp. 2-11, ACM Press/Springer-Verlag, January 1999

[9] Haidar, B., Joly, P., Haidar S., A Graph-based Approach to Automatically Chain Distributed Multimedia Indexing Services, in Proc. of 9th Int. Conf. on Internet and Multimedia Systems and Applications, ACTA Press, 2005;

[10] C.C. Hsu, W.W. Chu, and R.K. Taira, “A Knowledge-Based Approach for Retrieving Images by Content,” *IEEE Trans. Knowledge and Data Engineering*, vol. 8, no. 4, pp. 522-532, 1993

[11] JeongKyu Lee, Jung-Hwan Oh, and Sae Hwang. Strg-index: Spatio-temporal region graph indexing for large video databases. In SIGMOD Conference, pages 718–729, 2005

[12] Micarelli, A., Gasparetti, F., Sciarrone, F., Gauch, S., Personalized Search on the World Wide Web, Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web, LNCS 4321, Springer, 2007

[13] Micarelli, A., Sciarrone, F., Marinilli, M.: Web document modeling. In Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.), *The Adaptive Web*, LNCS. 4321, Springer, 2007

[14] Nack, F., Van Ossenbruggen, J., Hardman, L., "That Obscure Object of Desire: Multimedia Metadata on the Web (Part II)", In: IEEE Multimedia 12(1), pp. 54-63 January - March 2005

[15] A. Ono, M. Amano, M. Hakaridani, T. Satou, and M. Sakauchi, “A Flexible Content-Based Image Retrieval System with Combined Scene Description Keyword,” *Proc. Int’l Conf. Multimedia Computing and Systems*, pp. 201-208, 1996

[16] Salton, G., McGill, M.: Introduction to Modern Information Retrieval. McGraw-Hill (1983)

[17] Sarwar, B., Karypis, G., Konstan, J.A., Riedl, J.: Incremental SVD-Based Algorithms for Highly Scalable Recommender Systems. Proceedings of the Fifth International Conference on Computer and Information Technology, 2002

[18] Viola, P., Jones, M., Robust real-time object detection. Proceedings of Second International Workshop on Statistical and computational theories of vision, Canada *2001;*

[19] Yang, M. H., Kriegman, D., Ahuja, N., “Detecting faces in images: A Survey”, *IEEE Transaction Pattern Analysis and Machine Intelligence*, 2002