This paper is a pre-print version presented at the ISKO UK 2009 conference, 22-23 June, prior to peer review and editing. For published proceedings see special issue of Aslib Proceedings journal.

Exploiting Context awareness and annotation process to support multimedia information retrieval

Hanene Maghrebi [1] Amos David[2]
Hanene.maghrebi@loria.fr Amos.david@loria.fr

Team SITE-LORIA
Nancy Université
Laboratoire Lorrain de Recherche en Informatique et ses Applications
Campus Scientifique-BP 239-54506 Vandoeuvre les Nancy Cedex, France

Abstract

One of the characteristics of this age - information and knowledge age - is information overload and increase in multimedia information. Managing the increasing growth of multimedia information still poses some problems. The challenge is to propose relevant information to users among the large volume of multimedia information. Our approach consists in exploiting context awareness and annotation process to support multimedia information retrieval by supporting appropriate user-system interaction paradigms in order to better respond to user's multimedia information need.

Key words: multimedia information, context, annotation, indexing, user information needs

1. Introduction

One of the characteristics of this age - information and knowledge age - is information overload and increase in multimedia information. Multimedia information production is becoming a common practice among internet users as it could be noticed on YouTube[1] and Facebook[2]. On the contrary, there exists lot of challenges related to the management of increasingly growing multimedia information. One of the challenges is on how to retrieve relevant information among the large volume of multimedia information with respect to a given information problem.

Another problem in managing multimedia information is related to multi-component nature of multimedia information. Multimedia information is a composite of images, sound and texts expressed and disseminated using a particular media. It is a set of information involving at least two elements of these three. It draws its richness from the constituents of these three components. Multimedia information is complex considering the characteristics of its components and the type of relation between these components. These components are in their turn made up of sub-elements. Multimedia information is quite different from alphanumeric information in terms of composition and semantics. The complex structure of multimedia information demands a complex processing to retrieve the relevant information

[1] You tube (founded in 2005) is a video sharing website where users can upload, view and share video clips. http://www.youtube.com
[2] Face book is a social networking, founded in 2004. People can send message and share image and video

and to derive semantics and knowledge from its contents. In fact, in multimedia information, the text or the sound elements accompanying the image guides the user in his interpretation of the image and bring out a meaning which the user never thought of. “An image cannot be pinpointed with universal consensus because there is no consensus on universal meaning”. The semantic concepts depicted in, or otherwise emerging from, an image is specific to a user. When it is not obvious to grasp the meaning of image in the multimedia information, it becomes difficult to represent and manage it on the one hand and to express the information needs of the user on the other hand. As a matter of fact, user’s information needs can concern the full document or specific elements of information such as scene, shot in video, audio speech of a personality, etc. The challenge with multimedia information retrieval is to satisfy the different but precise user’s information needs. For instance a user’s need may concern a specific shot or image in a cinematographic document. This challenge can be split up into relevant questions:

- How can we represent multimedia information taking into consideration user’s needs?
- How can we bridge the semantic gap between user’s information needs and the information results proposed by the system?

In this paper, we review the challenges of multimedia information retrieval (MIR) in section 2. In section 3, we present a literature review on multimedia information indexing and representation. In the fourth section, we introduce our approach of exploiting context awareness and annotation process in supporting multimedia information retrieval, which is implemented in CO-ADMIRE (Context based multimedia information retrieval system), presented in section 5.

2. Challenges of multimedia information retrieval

Challenges of multimedia information retrieval are associated to the metadata representing multimedia content and to user’s expression of his information need. In multimedia information (image retrieval case), the user submits a query that is compared to the image on the system by their metadata. In most cases of image retrieval, a user employs semantic words to express his need. The image conveys denotative and connotative message. The denoted meaning of an image refers to the literal meaning given by the dictionary and can be understood by all users. For example the word “Watermelon” denotes a fruit. Connotation is a secondary signification added to the denotation. It refers to the associations that are connected to or suggested by the image. The connotation depends on the situation of the transmitter and the receiver of the image. “The connotative meanings of an image are generally based on the relationship between the image document and the image viewers in a particular situation” (Yon 2008). “Watermelon” can be connotated in the mind of the user as the holidays in Africa or an aliment sculptor. An image may generate different connotative meanings depending on image viewer’s context. Context can be cultural one as pointed out by semiotics essay (Barthes, 1977; O’Connor et al, 1999).

A user’s information need can be a connotative need, if the user is interested in semantic entities in multimedia information. The connotative need may have an implied meaning to the user. Connotative needs pose several problems as it is the impression and the sensation of the users that guide information needs. Such needs represented by keywords may be ambiguous hence making it difficult to retrieve relevant information from a multimedia information retrieval system.

The connotative meaning is one of the characteristics of image that are not easy to represent during indexing process. “Because of the notion that connotative attributes of an image are subject to an individual view’s interpretation, connotative messages tend to be ignored when developing image representation schemes or retrieval systems” (Yon 2008). For multimedia information, the expression of information need is subjective; there representation requires consideration of denoted and connoted aspect. A multimedia medium (e.g. video) enables us to see something (denoted aspect) and to feel or to imagine something else (connoted aspect). This polysemic aspect brings several meanings; the interpretation depends on the user; his culture, his background and the context in which the information is emitted and received.

The connotation aspect of Multimedia Information (MI) is often ignored in information retrieval system. While the image is represented using low level descriptors (texture, colour…), the user formulates his information needs according to the connoted aspect with semantic words. The matching between the information represented according to its denoted meaning and the user’s information needs having connoted meanings can only evoke a semantic gap.

The interpretation given by the user to the symbol of an image in MI is a problematic subject. Semantic and linguistic information associated to a perceived component of an image is not the same for all users. A non concordance between perceptual information on the image and the interpretation associated by various users can generate an ambiguity of interpretation. The user expresses his needs semantically whereas the representation of MI in the system is done in relation to the specificities of each medium.

3. Related work

There are existing approaches towards the development of multimedia information retrieval system (MIRS). These approaches were proposed with the aim of solving “semantic gap” problem dominant in multimedia information retrieval (MIR). Smeulders et al (2000) defined the semantic gap as “the lack of coincidence between the information that one can extract from the visual data and the interpretation that same data have for a user in a given situation”. The semantic gap is the distance between the user information need and the information proposed by the system.

Research works on MIRS are carried out using various approaches (Charhad et al 2005; Elgazel et al 2005). Information Systems, in general, have experienced some profound evolutions either in the objective of use – exploitation – or in the nature (typology) of managed information. The nature or the type of information that the information system manages, influences the functionalities of the system. Classical approach to multimedia information retrieval is to represent and index MI using the signal level (Valette 2001; Delmas 200; Elgazel et al 2005). There are two major ways of indexing using the classical approach:

- This first case of indexing is generally achieved automatically. Video and images are represented by their visual components e.g. colour, texture etc. The audio component is indexed by their frequency and structure. These descriptors known as “low level” are indexing component attributes. To find MI in response to user’s query, the system calculates the distances between the basic descriptors most similar to the image requested.
- The second case of indexing is structural level. The MI is organized in hierarchical order. The video, a fundamental component of MI, can be seen as sequences, scenes and shot (Charhad et al 2005; Maghrebi et David 2006). In this case the linear structure of the video does not permit access to semantic information

> which makes its use ambiguous. This calls for need to carry out a micro segmentation. For each plan it is a question of extracting a key image and segmenting a section of its area. Each area corresponds to a visual entity. Representation is done on these entities.

A criticism to the two indexing cases above concerns their low semantic level. This is because descriptors employed to represent the image do not account for the semantic aspect of the image. However, user-studies show that most users are interested in semantic entities rather than in visual appearance. The structural level and the low level can complement each other. The classical approach was followed by the semantic approach, which can be considered as a user approach. Hence the interest is now more on the user and his interpretation of the multimedia component than in the calculation of the distances between the image basic descriptors and user's query.

Recently annotation is used as technique for attempting to complete the "weakness" of MI indexing, in which keyword annotations are applied by users to label image (Enseret al 2005). Indeed, metadata (descriptors) in the indexing approaches above are destined to describe and model MI with an ultimate goal of responding to users' information needs. It is important to note however that if user's information needs are not integrated right from the conceptual stage of information retrieval system, it will be difficult to satisfy them after.

It will thus be interesting to represent user at the same level with document in order to better satisfy user's information needs in terms of time and result relevance. This poses another problem besides the semantic representation constraints of multimedia information which is a problem relating to the expression of user's information need.

4. Context –Aware retrieval

4.1 Problem relating to the expression of user's information need

User's information needs are complex to understand and to model or represent as it has been pointed out by (Wilson 1999-2000). An information need depicts "a lack in information" as regards a decisional problem. Users are always faced with the problem of clarifying or making explicit their information need. In fact, the capacity of a user to clarify his information requirements depends on his domain knowledge as well as his knowledge of the information system. In other words, knowledge of the area of search allows the user to better determine his information needs. The domain knowledge can be represented in information retrieval system in form of thesaurus and ontology. This kind of representation can help the user to specify his information needs.

User information needs satisfaction is tightly related to the representation of information on one hand and the integration of user information needs on the information retrieval system on the other hand. In order to study user information needs on multimedia information, we solicited a group of graduate and undergraduate student of the European Institute of Cinematography and Audiovisual (l'IECA). The students were administered a questionnaire designed to investigate three stages of multimedia information needs. At the first stage, students were asked to express their information needs, then to describe the kind of image expected to satisfy their information needs. We observed that the students find it difficult to express their information needs. We used semi-open questionnaire asking the students to provide description of how the information they are looking for is going to be used. We discovered that users find it easier to express the "use context" of the information they are looking for than expressing directly their information needs.

4.2 Introducing Context –Awareness in multimedia information retrieval process

User seeks information for a particular need. User's goals can be associated to the context of use of the information he is looking for. The satisfaction of user needs is related to the context in which the information will be used so we translate user's information need to his information use context. In our system, "information use context" is represented by four predefined attributes: teaching, learning, documentation and entertainment. We allow users to add other "information use context" attributes through annotation process to complete the representation of their information need. We call these attributes "open attributes" since every user can propose a new attributes related to his information use context if such attribute does not exist on the system. We can summarize our approach, on which we base our contextual multimedia information retrieval system, by Figure 1.

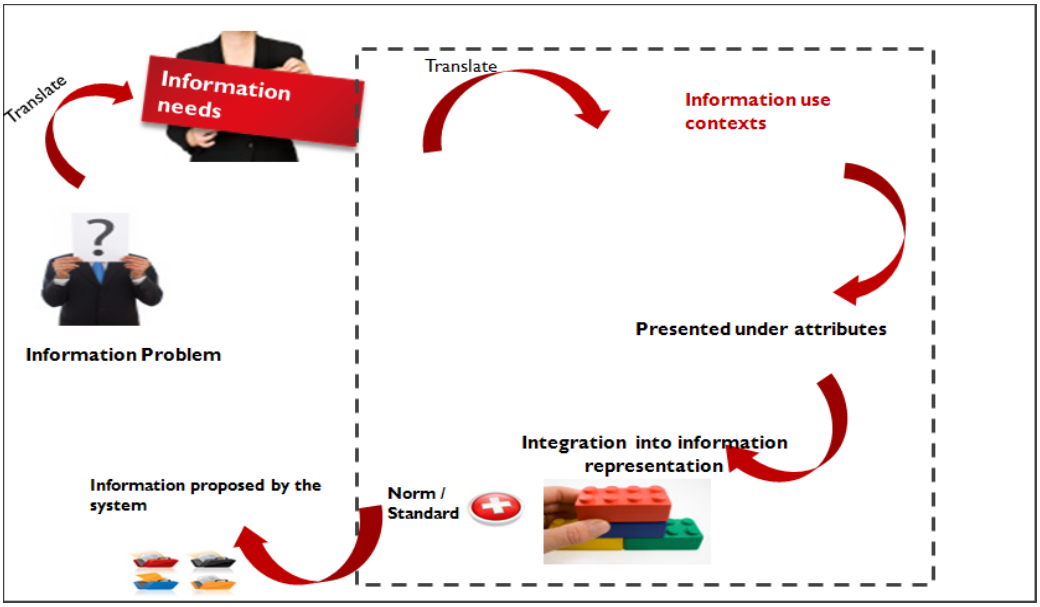

Figure 1: Integration process of user's information need into information retrieval system

We integrate other types of context to complete our context awareness approach. They include: the spatial context, temporal context, material context and cultural context. These contexts are added in form of attributes and values. Introducing Context-Awareness in multimedia information retrieval process is a form of personalization of the interaction between the user and the system since context awareness can be employed to precise the information that the system propose as response to user's specific query. Context can be exploited to refine user's query. Its exploitation can also be in storing document and sub component under different context.

4.3 Context – document storage

We represented and indexed multimedia information and its subcomponents in a best possible detailed form using inspired attributes (metadata) from Dublin Core and organizational structure of cinematographic document norms but this representation does not on its own resolve the semantic problem hence we associate each document entity to use contexts as could be seen in figure 2. We allow user to participate in the indexing process through annotation by adding relevant use contexts to a consulted document.

Integrating user in the indexing process can produce a "vocabulary documentary" which can enrich MI indexing. A verification of vocabularies in return is necessary to avoid polysemic and synonymic problem. User's annotation can constitute a complement to the traditional metadata plan.

Existing context-awareness retrieval systems provide different ways to retrieving documents using "time context" or "location context" but to the best of our knowledge, none of such systems supports document retrieval based on information use contexts.

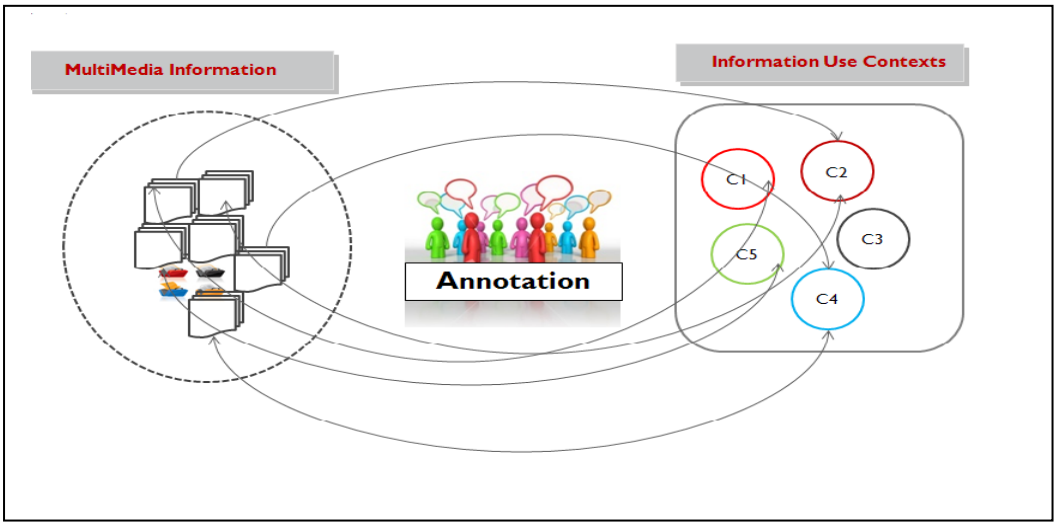

Figure 2: Associating multimedia document and its subcomponents to different information use contexts through annotation process

5. Co-Admire: Contextual system for multimedia information retrieval

Co-Admire (COntext bAsed MultimeDia Information Retrieval SystEm) is an open environment for user to clarify his information problem, to define his information need, to translate the demand of information into queries and to specify the context of use of the retrieved information. The open system created is a contextual system that can adapt his behavior to user's information need and to the context of the information problem.

By Open System (figure3), we imply a system made by user for user. A system where users have the possibility to propose or to add what they think is missing as attributes or sub-

attributes (in other words the metadata of retrieval) and to annotate the existing attributes as well as data. We share the same opinion with Zacklad 2007 on open system concept. He sees an open system as not only a tool for accessing relevant documents or some relevant documentary fragments but also as a tool that delimit the problem contours by identifying the existing information and those that do not exist. We see open information system approach as combining personalization in IR with user's participation in indexing and representation of multimedia documents to resolve the semantic problem related to MIR.

In Co-admire; a user can carry out a simple search for documents. He can also search by cross-linking two or three attributes representing the documents in the base. He may cross-link the context attributes with the document attributes. This cross linking of attributes also called concurrence querying allows user to build knowledge on existing documents in the database as well as on the use context of these types of documents.

The system keeps track of the search history of users with the goal of enhancing knowledge sharing among them. It is in knowledge sharing perspective that information indexing and retrieving become interesting. This allows a user to have an idea of the types of information problem that can be solved using the system and to acquire knowledge on formulating and representing his problem. The user can also see the annotation and evaluation on the results proposed by the system to the queries of other users who had used the system in the past. He can also see their information needs and use context. Indexing can then take a new scope which can be defined under collaborative work between different users.

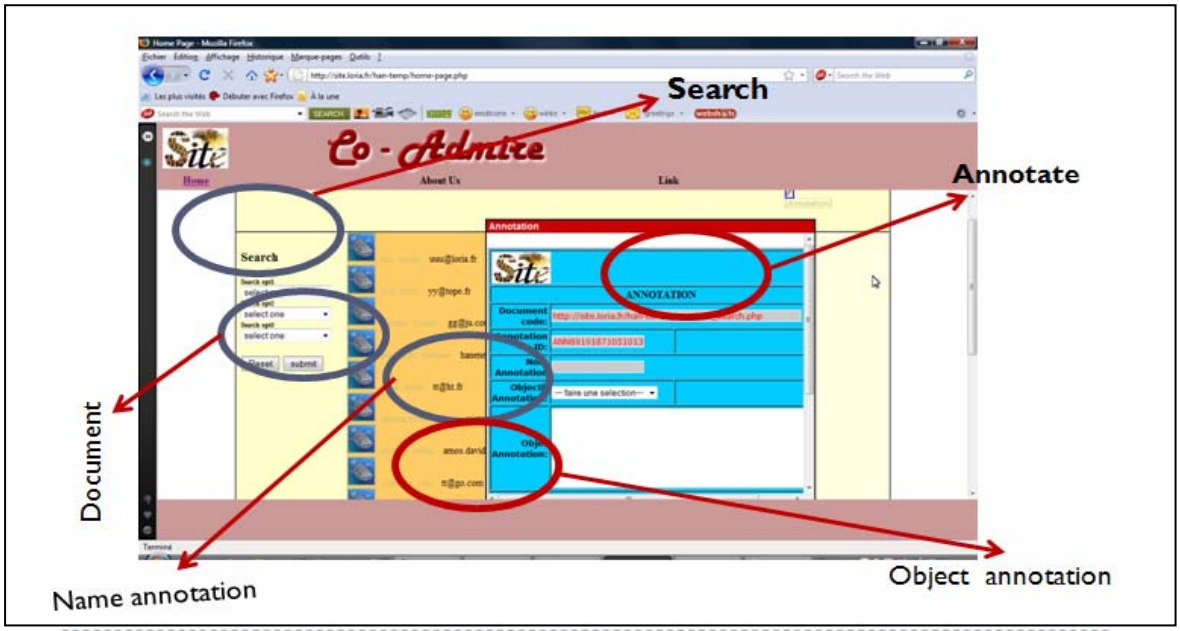

Figure 4: screen copy of Co-Admire: Contextual system for multimedia information retrieval

6. Conclusion

Our research application is related to cinematographic corpus. These are the collections of films proposed by the European Institute of Cinematography and Audiovisual (IECA). Our potentials users are lecturers, multimedia producers and socio-economic actors. The

information needs for all these users are not the same, the information use context are also not the same. A single multimedia document (e.g. a movie) can be associated to many information use context. Such a document may be needed by different user for different use context. We represented user's information needs at the same level as multimedia information using information use context. We allow user to participate in the indexing process through annotation. We can conclude that integrating use context into multimedia information retrieval will improve retrieval precision and lead to better retrieval speed. Our approach was implemented in Co-admire. The future work would be to experiment our multimedia information system and evaluate the effectiveness and usability of the system.

References

Barthes, R. (1964), "Rhetoric of the image''. Available at: http://www.oasisfle.com/doc_pdf/roland_barthes_rhetorique_image.pdf (accessed 12 March 2009)

Charhad M, Zrigui M, Quénot G, (2005), Une approche conceptuelle pour la modélisation et la structuration sémantique des documents vidéo'', SETIT, tunisia. Available at: http://www.setit.rnu.tn/last_edition/setit2005/applications/343.pdf (accessed March 2009)

Delmas,P. (2000),''Extraction des contours de lèvres d'un visage parlant par contours actifs, Application à la communication multimodale'' *PhD Thesis, National Polytechnic Institute, Grenoble, France*.

Elghazel, A. et al, (2005), "Approche textuelle pour la recherche d'image'', SETIT, Tunisie, 2005, Available at: http://liris.cnrs.fr/Documents/Liris-1953.pdf (accessed March 2009)

Egyed –Zsigmond E. (2003), "Gestion de connaissance dans une base de documents multimédias'', thèse en informatique et information pour la société, *PhD Thesis, INSA, Lyon*.

Enser, P.G.B. Sandom, C.J., Lewis, P.H. (2005), Automatic annotation of images from the practitioner perspective. In Leow, W.K., Lew, M.S., Chua, T.S., Ma, W.Y., Chaisorn, L., Bakker, E.M., eds.: CIVR. Volume 3568 of Lecture Notes in Computer Science., Singapore, Springer 497–506
Available at: http://www.springerlink.com/content/47gkrekby8h55569/fulltext.pdf (accessed March 2009)

Maghrebi, H. and David, A. (2006) "toward a model for the representation of multimedia information based on users' needs: economic intelligence approach", proceeding of 4th International conference on Multimedia and Information and Communication Technologies in Education m-ICTE (proceedings), 195-200. Available at: http://www.formatex.org/micte2006/pdf/195-200.pdf (accessed March 2009)

Maghrebi, H. and David, A. (2007), "Integrating users' needs into multimedia information retrieval system". Proceeding of 3rd International Conference on Computer Science ATINER, Greece. Retrieved March, 2009, Available at: http://hal.inria.fr/inria-00167511/fr/ (accessed March 2009)

Marchionini, G. (2004), "From information retrieval to information interaction," proceeding of the 26th European Conference on Information Retrieval research, ECIR 2004, Sunderland, UK(p1- 11). Available at: http://www.ils.unc.edu/~march/ECIR.pdf (accessed March 2009)

O'Connor, B. et al. (1999), "user reactions as access mechanism: an exploration based on captions of image". *Journal of the American society for information science*, 50,681-697

Valette, S. (2001), Modèles de maillages déformables 2D et multi-résolution surfaciques 3D sur une base d'ondelettes., *PhD Thesis, INSA Lyon*, France.

Wilson, T. (1999), ''Models in information behaviour research''. *Journal of documentation*, 55(3), P. 249-269.

Available at: http://informationr.net/tdw/publ/papers/1999JDoc.html (accessed March 2009)

Wilson, T. (2000), ‘‘recent trend in user studies: action research and qualitative methods”. *Information research* Available at: http://informationr.net/ir/5-3/paper76.html (accessed March 2009)

Smeulders, A. et al,(2000), “Content-Based Image Retrieval at the End of the Early Years,” proceeding of IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 12, Dec. 2000, pp. 1349-1380.
Available at: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=00895972 (accessed 30 March 2009)

Yon, J. (2008), “Searching for image conveying connotative meanings: an exploratory cross-cultural study”, *library and information science Research*, Vol. 30, Issue 4, pp. 312-318.

Zacklad, M. (2007), “Classification, thesaurus, ontologies, folksonomies: A comparison from the Open Information Research perspective”. In Information Sharing in a Fragmented World: Crossing Boundaries, proceeding of 35 th Canadian Association for Information Science. Available at: http://www.cais-acsi.ca/proceedings/2007/zacklad_2007.pdf (accessed 24March 2009)