UNIVERSITY
OF TRENTO

DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.disi.unitn.it

GLOCAL MULTIMEDIA RETRIEVAL

Giulia Boato, Claudio Fontanari, Fausto Giunchiglia,
and Francesco G.B. De Natale

January 2008

Technical Report # DISI-09-002

GLOCAL MULTIMEDIA RETRIEVAL

Giulia Boato, Claudio Fontanari, Fausto Giunchiglia, and Francesco G.B. De Natale*

Department of Information Engineering and Computer Science
*Department of Mathematics
University of Trento
Via Sommarive 14, 38100 Trento, Italy
{giulia.boato, claudio.fontanari, fausto.giunchiglia, francesco.denatale}@unitn.it

ABSTRACT

Personal experience is intrinsically local while common knowledge is global. As a consequence, standard multimedia search engines suffer from a gap between local content and global concept, due to the diversity of context. Here we design a new multimedia retrieval system which integrates local diversity into an evolving global knowledge. A suitable personalization of global ontologies allows to extract information from multimedia content according to context-sensitive relevance, while a peer-to-peer communication model provides a decentralized and scalable logical architecture for media search.

***Index Terms*—** Multimedia retrieval, Semantic gap, Ontology, Event, P2P search

1. INTRODUCTION

Retrieval of multimedia data is still a widely open problem despite its relevance. In particular, the intrinsic limitation of purely text-based search has led in the last decade to a growing interest in content-based image retrieval (CBIR). Indeed, even through textual search engines are still the most widely used, CBIR is now a mature discipline. For a detailed overview of this research area we refer to the classical exposition [1] and to its recent update [2]. As it is well known, the crucial challenge of CBIR is the so-called semantic gap, formalized in [1] as the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. However, we point out that the current availability of more and more devices that allow users to generate new multimedia content, by capturing their own experience in images and videos, mixing it with digital material collected from the web, and finally sharing it with other users, highlights novel aspects of such a gap between low-level features and high-level semantics. In particular, a local-to-global dynamic arises, introducing a *glocal gap* between local content and global concept. Indeed, in our opinion this is just a facet of the more general gap between experience and knowledge, which has been a long-standing challenge addressed by theoretical philosophy from Plato to Kant. From this point of view, it turns out that matching content and concept is a priori impossible without keeping context into account. After the pioneering contribution by [3], the ICT community is becoming aware that diversity is a must, not only as the unavoidable reason for diverging viewpoints and conflicts, but also as an essential feature to be preserved and exploited in order to develop better technology (diversity-aware classification, navigation and search). Summing up, the situation is as in Figure 1.

Figure 1

To be more concrete, let us introduce a motivating example. By definition, a news is a fact happening in a specific place at a certain time, relying on a local context but with a global relevance. Such a local-to-global dynamic has been emphasized in the last few years by the spreading of multimedia devices allowing to capture and share news touching personal experience. In this scenario, the role of the reporter is consequently changing from the only producer of professional multimedia content to the collector of many different sources capturing also the experience of directly involved people. For instance, nowadays reports on bad weather emergency share with the general audience homemade pictures and videos witnessing several viewpoints and contexts. This specific example will allow us to illustrate a new framework for media search enabling

both personalization of global knowledge and sharing of local experience.
In the following we propose our approach to bridge the glocal gap by organizing knowledge into ontologies and experience around events and we design a distributed search framework based on a logical P2P architecture. In both cases, we are going to discuss the state-of-the-art of the exploited tools and to outline the novelties of our integrated glocal approach.

2. THE SOLUTION

In order to close the aforementioned glocal gap, a bridge between knowledge and experience as captured by media is needed. On the semantic side, common knowledge can be formalized into ontologies, which are explicit specifications of conceptualizations [4]. Ontologies are often thought of as directed graphs whose nodes represent concepts and whose edges represent relation between concepts. The backbone structure of the ontology graph is a taxonomy in which all relations are "is-a" relations, whereas the remaining structure of the graph supplies auxiliary information about the modelled domain. Following the definition given in [5], a (formal) lightweight ontology is an ontology which can be represented in the form of a rooted tree in which nodes stand for concepts and edges stand for the subsumption relation between the two concepts they connect. In their simplest version, one can think of lightweight ontologies as ontologies consisting of backbone taxonomies only.
The idea of applying ontologies to retrieval of multimedia content is already present in the literature (see in particular [6], [7], and the book [8]) and it is clear that they can provide an enriched description of the semantics involved in the retrieval process. However, to make such an intuition into an effective tool we propose to construct ontologies following a bottom-up instance-driven approach inspired by library science. From this discipline we borrow the facets theory, as originally envisaged by Ranganathan in [9] as a promising input to the creation of our knowledge model. Other examples of large-scale ontologies constructed in a bottom-up (data-driven) fashion are provided by Yago [10] and KnowItAll [11]. Next, we apply lightweight ontologies to multimedia content according to an event-based model. Indeed, events can be viewed as elementary logic units of experience where content is enriched by concept (see Figure 2).

Figure 2

Formally, an event is defined as a significant occurrence or happening at a single point in space-time. In fact, events can be viewed as objects with time and space as their primary attributes [12]. As the same author points out in the recent contribution [13], a common event model has not already been developed, despite ubiquity of events in multimedia.
We stress that the notion of event is intrinsically local and heavily relies on context, hence it can properly describe experience and bridge the gap between local content and local concept. On the other hand, to produce general knowledge and bridge the glocal gap described in Figure 1, a global ontology should come into play. More precisely, the local picture needs to be enriched by a bidirectional link with such a global ontology, in order to achieve both personalization of global knowledge and sharing of local experience. In particular, contextualization of common knowledge is obtained via local CRUD (pruning and refining) and by grounding concepts on context, thus defining a *glocal ontology* (see Figure 3).

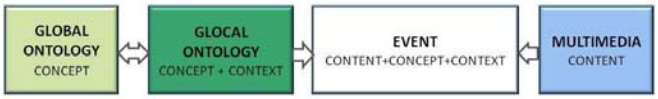

Figure 3

Turning to the promised motivating example, as a global ontology for bad weather we consider the toy lightweight ontology sketched in Figure 4. In the local context of two different places, such as Venice and the Dolomites, the procedures of pruning and refining produce the glocal ontologies depicted in Figure 5.

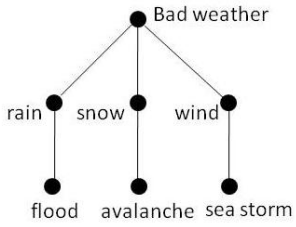

Figure 4

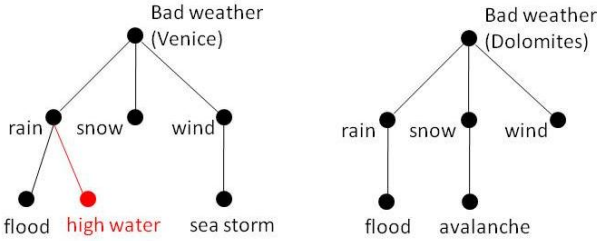

Figure 5

From our point of view, a query like "bad weather in Venice" is matched with the corresponding multimedia content by refining the global ontology and inserting the local context-dependent concept of "high water". The result shall be the retrieval of a collection of multimedia data capturing experience into events. It is now clear that the event is the locus where the glocal gap is closed by exploiting local diversity (see Figure 6).

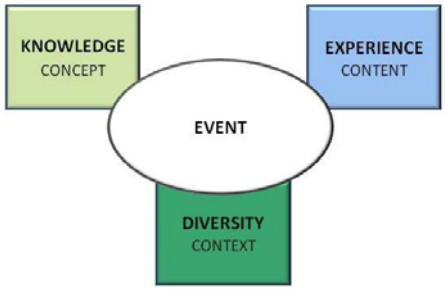

Figure 6

Notice that the proposed bridge on the glocal gap requires just one global ontology (managed by professional experts as in library science) and many glocal ontologies corresponding to different users. Hence in our view global multimedia search is naturally distributed into a P2P local search according to a logical architecture where the global ontology acts as a super-peer (see Figure 7).
This approach is consistent with the current paradigm shift from centralized to P2P search, ongoing also in the multimedia retrieval community (see for instance [14]). In our model, queries are redirected by the global supervisor to the corresponding nodes and locally translated into a content-based context-aware search.

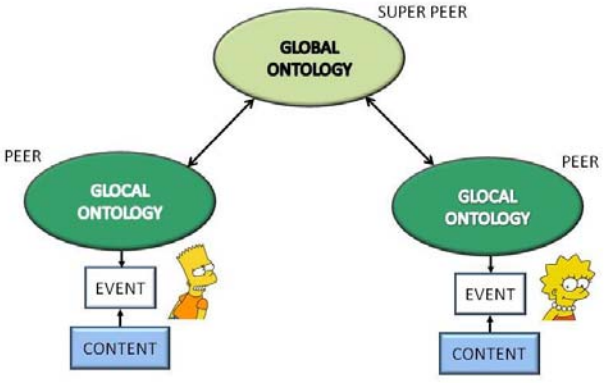

Figure 7

3. CONCLUSION

The semantic gap between low-level features and high-level semantics affecting current CBIR paradigm has been rephrased as a glocal gap between local content and global concept, and then addressed via both personalization of global knowledge into glocal ontologies and organization of local experience around events. As a consequence, we have designed a decentralized and scalable logical architecture for media search to extract information from multimedia content according to context-sensitive relevance.
Even though the outlined picture is comprehensive and consistent, several research issues are still open and are the subject of current investigation. In particular, we are aware that the dynamic enriching of global ontology by generalization of local knowledge is still a challenging problem.

4. REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta and R. Jain. Content-Based Image Retrieval at the End of the Early Years. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, Issue 12, 1349-1380, 2000.

[2] R. Datta, D. Joshi, J. Li and J.Z. Wang. Image Retrieval: Ideas, Influences, and Trends of the New Age. ACM Computing Surveys, Volume 40, Issue 2, 5:1-5:22, 2008.

[3] F. Giunchiglia. Managing Diversity in Knowledge. Keynote talk, European Conference on Artificial Intelligence (ECAI-06), 2006.

[4] T. R. Gruber. A translation approach to portable ontology specifications. Knowl. Acquis. 5(2),199–220, 1993.

[5] F. Giunchiglia and I. Zaihrayeu. Lightweight Ontologies. To appear in "The Encyclopedia of Database Systems", Springer, 2008.

[6] V. Mezaris, I. Kompatsiaris, N.V. Boulgouris, M.G. Strintzis, Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval. IEEE Transactions on Circuits and Systems for Video Technology, Volume 14, Issue 5, 606-621, 2004.

[7] D. Vallet, P. Castells, M. Fernandez, P. Mylonas and Y. Avrithis. Personalized Content Retrieval in Context Using Ontological Knowledge. IEEE Transactions on Circuits and Systems for Video Technology. Volume 17, Issue 3, 336-346, 2007.

[8] Y. Kompatsiaris and P. Hobson. Semantic Multimedia and Ontologies. Theory and Applications. Springer 2008.

[9] S. R. Ranganathan. Prolegomena to Library Classification (Classic). (Ranganathan Series in Library Science, 20), Asia Publishing House, London, 1967.

[10] F. Suchanek et al. YAGO: A Core of Semantic Knowledge. In WWW, 2007.

[11] O. Etzioni, M. J. Cafarella, D. Downey, Ana-Maria Popescu, Tal Shaked, S. Soderland, D. S. Weld and A. Yates. Unsupervised named-entity extraction from the Web: An experimental study. Artif. Intell. 165(1), 91-134, 2005.

[12] R. Jain. Experiential computing. Communications of the ACM. Volume 46, Issue 7, 48 – 55, 2003.

[13] U. Westermann and R. Jain. Toward a Common Event Model for Multimedia Applications. IEEE Multimedia. Volume 14, Issue 1, 19-29, 2007.

[14] A. Vlachou, C. Doulkeridis, D. Mavroeidis and M. Vazirgiannis. Designing a Peer-to-Peer Architecture for Distributed Image Retrieval. Adaptive Multimedial Retrieval: Retrieval, User, and Semantics. LNCS 4918, 182-195, 2008.