SAPIR: Scalable and Distributed Image Searching

Fabrizio Falchi*, Mouna Kacimi†, Yosi Mass‡, Fausto Rabitti*, Pavel Zezula§

* ISTI-CNR - Pisa, Italy - Email: {fabrizio.falchi|fausto.rabitti}@isti.cnr.it
† Max-Planck-Institut für Informatik - Saarbrücken, Germany - Email: mkacimi@mpi-inf.mpg.de
‡ IBM Haifa Research Lab - Haifa, Israel - Email: yosimass@il.ibm.com
§ Masaryk University - Brno, Czech Republic - Email: zezula@fi.muni.cz

***Abstract*—In this paper we present a scalable and distributed system for image retrieval based on visual features and annotated text. This system is the core of the SAPIR project. Its architecture makes use of Peer-to-Peer networks to achieve scalability and efficiency allowing the management of huge amount of data.**

For the presented demo we use 10 million images and accompanying text (tags, comments, etc.) taken from Flickr. Through the web interface it is possible to efficient perform content-based similarity search, as well as traditional text search on the metadata annotated by the Flickr community. Fast complex query processing is also possible combining visual features and text.

We show that the combination of content-based and text search on a large scale can dramatically improve the capability of a multimedia search system to answer the users needs and that the Peer-to-Peer based architecture can cope with the scalability issues (response time obtained for this demo over 10 million images is always below 500 milliseconds).

***Index Terms*—Peer-to-Peer, metric spaces, distributed, scalability, MPEG-7, similarity search.**

I. Introduction

Non-text data, such as images, music, animations, and videos is nowadays a large component of the Web. However, web tools for performing image searching (such the ones provided by Google, Yahoo!, or MSN Live Search) simply index the text associated with the image. Web search is dominated today by text only indexes enriched by page rank algorithms, thus search for audio-visual content, it is limited to associated text and metadata annotations.

Image indexing methods based on content-based analysis or pattern matching (which typically analyze the characteristics of images, i.e., features, such as colors and shapes) are usually not exploited at all. In fact, for this kind of data the appropriate search methods are based on similarity paradigms that typically exploits range queries and nearest neighbor queries. These queries are computationally more intensive than text search, because conventional inverted indexes used for text are not applicable for such data.

The European project SAPIR (Search on Audio-visual content using Peer-to-peer Information Retrieval)[1] aims at breaking this technological barrier by developing a large-scale, distributed Peer-to-Peer infrastructure that will make it possible to search for audio-visual content by querying the specific characteristics (i.e., features) of the content. SAPIR's goal is to establish a giant Peer-to-Peer network, where users are peers that produce audiovisual content using multiple devices (e.g., cell phones) and service providers will use more powerful peers that maintain indexes and provide search capabilities

"A picture is worth a thousand words" so using an image taken by a cell phone to find information about e.g. a monument we bump into or singing a melody as a search hint for a full song, combined with optional metadata annotations and user and social networking context will provide the next level of search capabilities and precision of retrieved results.

[1] http://www.sapir.eu/

II. SAPIR Architecture

Although many similarity search approaches have been proposed, the most generic one considers the mathematical metric space as a suitable abstraction of similarity [1]. The simple but powerful concept of the metric space consists of a domain of objects and a metric distance function that measures the proximity of pairs of objects. A distance, to be a metric, must satisfy a set of simple constrains the most important of which is the triangle inequality.

The metric space approach has been proved to be very important for building efficient indexes for content based similarity searching. A survey of existing approaches for centralized structures (e.g. M-tree, can be found in [1]).

However, searching on the level of features exhibits linear scalability with respect to the data search size. The reason is that for this kind of data the appropriate search methods are based on similarity paradigms that typically exploits range queries and nearest neighbor queries which very much intensive because conventional inverted indexes used for text are not applicable for such data.

Very recently scalable and distributed index structures based on Peer-to-Peer networks have also been proposed for similarity searching in metric spaces and are used in the context of the SAPIR project - i.e. GHT* [2], VPT* [3], MCAN [4], M-Chord [5] (see [3] These index structures have been proved to provide scalability for similarity search adding resources as the dataset grows. Peer-to-Peer architectures are convenient approach and a common characteristic of all these existing approaches is the autonomy of the peers with no need of central coordination or flooding strategies. Since there are no bottlenecks, the structures are scalable and high performance is achieved through parallel query execution on individual peers.

In SAPIR also text will be indexed using a Peer-to-Peer architecture called MINERVA [6]. In MINERVA each peer is considered autonomous and has its own local search engine with a crawler and a local index. Posting meta-information into the Peer-to-Peer network the peers share their local indexes. This meta-information contains compact statistics and quality-of-service information, and effectively forms a global directory. The Peer-to-Peer engine uses the global directory to identify candidate peers that are most likely to provide good query results. More information about MINERVA can be found in [6].

An IR-style query language for multimedia content based retrieval has been developed for SAPIR. It exploits the XML representation of MPEG-7 and it is an extension of the XML Fragments query language that was originally designed as a Query-By-Example for text-only XML collections. Detailed information can be found in [7].

In SAPIR it is also possible to perform complex similarity search combining result lists obtained using distinct features, GPS information and text. To this aim, state of the art algorithms for combining results are used (e.g., [8]). For multi-feature indexing SAPIR makes uses of the MUFIN (Multi-feature Indexing Network) which is built over the MESSIF [2] (Metric Similarity Search Implementation Framework) architecture [9].

The web user interface used for this demo is derived from the one we developed for a photo album application built upon MILOS [10] (a centralized Multimedia Content Management System).

III. Dataset

For the presented demo the dataset consists of 10 million images taken from Flickr [3]. Each image has metadata information annotated by the users community (e.g., tags, location, comments, etc.).

To perform content-based image retrieval we make use of 5 MPEG-7 Visual Descriptors - ScalableColor, ColorStructure, DominantColor, EdgeHistogram and HomogeneousTexture. The extraction of these features from the images typically requests from 1 to 2 seconds on a nowadays standard PC. Thus, to process tens of millions images a distributed environment was required. We decided to use a Grid infrastructure. In particular we used the DILIGENT [4] (A testbed DIgital Library Infrastructure on Grid ENabled Technology) project which delivers a Grid production infrastructure shared by a large number of European organisations on the EGEE [5] (Enabling Grids for E-sciencE) project.

Overall 44 thousand jobs were successfully executed on the grid, processing around 37 million images. This generated approximately 112 million text and images objects (4,55 TB of data) that contain more then 150 million extracted features. The target for the project is 100 million images. More detailed information (including job distribution per site) can be found in [11].

It is so far the largest test bed of multimedia content (it will grow up to 100 million Flickr's images) available not only inside SAPIR project but also, in the near future, to the research community.

[2] http://lsd.fi.muni.cz/trac/messif

[3] http://www.flickr.com

[4] http://www.diligentproject.org

[5] http://www.eu-egee.org

Acknowledgments

This work has been partially supported by the SAPIR (Search In Audio Visual Content Using Peer-to-Peer IR) project, funded by the European Commission under IST FP6 (Sixth Framework Programme, Contract no. 45128).

The development and preparation of the demo has involved a large number of people from several SAPIR project partners. In particular we would like to mention: Michal Shmueli-Scheuer and Benjamin Sznajder from IBM Haifa Research Lab; Paolo Bolettieri, Claudio Gennaro, Claudio Lucchese, Matteo Mordacchini, Raffaele Perego and Tommaso Piccioli from ISTI-CNR; Michal Batko, Vlastislav Dohnal, David Novak and Jan Sedmidubsky from Masaryk University; Tom Crecelius from Max-Planck-Institut für Informatik.

References

[1] P. Zezula, G. Amato, V. Dohnal, and M. Batko, *Similarity Search. The Metric Space Approach*, ser. Advances in Database Systems. Springer Science + Business Media, Inc., 2006, vol. 32.

[2] M. Batko, C. Gennaro, and P. Zezula, "Similarity grid for searching in metric spaces." in *Peer-to-Peer, Grid, and Service-Orientation in Digital Library Architectures. 6th Thematic Workshop of the EU Network of Excellence DELOS, Revised Selected Papers*, ser. LNCS, vol. 3664. Springer-Verlag Berlin Heidelberg, 2004, pp. 25–44.

[3] M. Batko, D. Novak, F. Falchi, and P. Zezula, "On scalability of the similarity search in the world of peers," in *InfoScale '06: Proceedings of the 1st international conference on Scalable information systems*. New York, NY, USA: ACM Press, 2006, p. 20.

[4] F. Falchi, C. Gennaro, and P. Zezula, "A content-addressable network for similarity search in metric spaces," in *DBISP2P '05: Proceedings of the the 2nd International Workshop on Databases, Information Systems and Peer-to-Peer Computing, Trondheim, Norway*, ser. Lecture Notes in Computer Science, vol. 4125. Springer, 2005, pp. 98–110.

[5] D. Novak and P. Zezula, "M-chord: a scalable distributed similarity search structure," in *InfoScale '06: Proceedings of the 1st international conference on Scalable information systems*. New York, NY, USA: ACM Press, 2006, p. 19.

[6] M. Bender, S. Michel, P. Triantafillou, G. Weikum, and C. Zimmer, "MINERVA: Collaborative P2P Search," in *VLDB '05: Proceedings of the 31st international conference on Very large data bases*. VLDB Endowment, 2005, pp. 1263–1266.

[7] J. Mamou, Y. Mass, M. Shmueli-Sheuer, and B. Sznajder, "Query language for multimedia content," in *Procedding of the Multimedia Information Retrieval workshop held in conjunction with ACM SIGIR Conference - 27 July 2007, Amsterdam*, 2007.

[8] R. Fagin, A. Lotem, and M. Naor, "Optimal Aggregation Algorithms for Middleware," *CoRR*, vol. cs.DB/0204046, 2002. [Online]. Available: http://arxiv.org/abs/cs.DB/0204046

[9] M. Batko, D. Novak, and P. Zezula, "Messif: Metric similarity search implementation framework," in *DELOS Conference 2007: Working Notes, Pisa, 13-14 February 2007*. Information Society Technologies, 2007, pp. 11–23.

[10] G. Amato, P. Bolettieri, F. Debole, F. Falchi, F. Rabitti, and P. Savino, "Using milos to build a multimedia digital library application: The photobook experience." in *Research and Advanced Technology for Digital Libraries, ECDL 2006*, ser. LNCS, vol. 4172. Springer-Verlag Berlin Heidelberg, 2006, pp. 379–390.

[11] "Diligent data challenges," 2007, [Online; accessed 12-October-2007]. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/DILIGENT/DiligentFlickrDC