UPGRADE is the European Journal for the Informatics Professional, published bimonthly at <http://www.upgrade-cepis.org/>

Publisher
UPGRADE is published on behalf of CEPIS (Council of European Professional Informatics Societies, <http://www.cepis.org/>) by **Novática** <http://www.ati.es/novatica/>, journal of the Spanish CEPIS society ATI (*Asociación de Técnicos de Informática,* <http://www.ati.es/>)

UPGRADE monographs are also published in Spanish (full version printed; summary, abstracts and some articles online) by **Novática**

UPGRADE was created in October 2000 by CEPIS and was first published by **Novática** and **INFORMATIK/INFORMATIQUE**, bimonthly journal of SVI/FSI (Swiss Federation of Professional Informatics Societies, <http://www.svifsi.ch/>)

UPGRADE is the anchor point for **UPENET** (UPGRADE European NETwork), the network of CEPIS member societies' publications, that currently includes the following ones:
- **Informatik-Spektrum**, journal published by Springer Verlag on behalf of the CEPIS societies GI, Germany, and SI, Switzerland
- **ITNOW**, magazine published by Oxford University Press on behalf of the British CEPIS society BCS
- **Mondo Digitale**, digital journal from the Italian CEPIS society AICA
- **Novática**, journal from the Spanish CEPIS society ATI
- **OCG Journal**, journal from the Austrian CEPIS society OCG
- **Pliroforiki**, journal from the Cyprus CEPIS society CCS
- **Pro Dialog**, journal from the Polish CEPIS society PTI-PIPS

Editorial correspondence: Llorenç Pagés-Casas <pages@ati.es>
Advertising correspondence: <novatica@ati.es>

UPGRADE **Newslist** available at
<http://www.upgrade-cepis.org/pages/editinfo.html#newslist>

ISSN 1684-5285

Monograph of next issue (April 2007)
"Information Technologies for Visually Impaired People"
(The full schedule of **UP**GRADE is available at our website)

CEPIS **UPGRADE**
The European Journal for the Informatics Professional
http://www.upgrade-cepis.org
Vol. VIII, issue No. 1, February 2007

Monograph: Next Generation Web Search
(published jointly with Novática*)
Guest Editors: *Ricardo Baeza-Yates, José-María Gómez-Hidalgo, and Paolo Boldi*

2 Presentation. The Future of Web Search — *Ricardo Baeza-Yates, Paolo Boldi, and José-María Gómez-Hidalgo*

5 Efficient Sparse Linear System Solution of the PageRank Problem — *Gianna M. Del Corso, Antonio Gullì, and Francesco Romani*

12 Learning to Analyze Natural Language Texts — *Giuseppe Attardi*

19 SNAKET: A Personalized Search-result Clustering Engine — *Paolo Ferragina and Antonio Gullì*

27 The Multimodal Nature of the Web: New Trends in Information Access — *Luis-Alfonso Ureña-López, Manuel-Carlos Díaz-Galiano, Arturo Montejo-Raez, and Mª Teresa Martín-Valdivia*

33 Adversarial Information Retrieval in the Web — *Ricardo Baeza-Yates, Paolo Boldi, and José-María Gómez-Hidalgo*

41 GERINDO: Managing and Retrieving Information in Large Document Collections — *Nivio Ziviani, Alberto H. F. Laender, Edleno Silva de Moura, Altigran Soares da Silva, Carlos A. Heuser, and Wagner Meira Jr.*

49 Research Directions in Terrier: a Search Engine for Advanced Retrieval on the Web — *Iadh Ounis, Christina Lioma, Craig Macdonald, and Vassilis Plachouras*

57 Yahoo! Research Barcelona: Web Retrieval and Mining — *The Yahoo! Research Team*

UPENET (UPGRADE European NETwork)

59 From **Novática** (ATI, Spain)
Informatics Profession
The Maturity of IT Professionalism in Europe — *Sean Brady*

68 From **Pro Dialog** (PTI-PIPS, Poland)
Graphical Interfaces
Portable Declarative Format for Specifying Graphical User Interfaces — *Zbigniew Fryźlewicz and Rafał Gierusz*

75 From **Novática** (ATI, Spain)
Next-generation Web
Blogs: On the Cutting Edge of the Next-generation Web — *Antonio Miguel Fumero-Reverón and Fernando Sáez-Vacas*

CEPIS NEWS

83 Harmonise Project: Building up to the Final Report—*François-Philippe Draguet*

84 News & Events: European Funded Projects and News Updates

* This monograph will be also published in Spanish (full version printed; summary, abstracts, and some articles online) by **Novática**, journal of the Spanish CEPIS society ATI (*Asociación de Técnicos de Informática*) at <http://www.ati.es/novatica/>.

The Multimodal Nature of the Web: New Trends in Information Access

Luis-Alfonso Ureña-López, Manuel-Carlos Díaz-Galiano, Arturo Montejo-Raez, and Mª Teresa Martín-Valdivia

The rapid evolution of the World Wide Web has changed our view of it. It has turned into a collaborative framework where technological and social trendstrends come together, resulting in the over exploited term Web 2.0. In this new multimodal and multilingual paradigm, all our techniques for the search and retrieval of information need to be applied, managing not only textual information, but also visual data (images or videos) that can help to improve our systems. In the present paper, along with a brief analysis of the described scenario, we introduce an experience in the medical domain for the retrieval of multimodal information (text and images).

Keywords: CBIR, Content-Based Image Retrieval, Medical Information, Multilingual Information, Multimodal Information Retrieval, Web 2.0.

1 Introduction

The World Wide Web has changed substantially since its beginningss in 1990. In fact we are currently witnesses to a revolution of the concept of what the web is [4]. This is a crucial stage in the development of the so-called knowledge society, and one which is also determined by the intersection of social and technological trends in what has been termed "Web 2.0". The social trend is the removal of barriers between information providers and consumers, the exploitation of the collective intelligence as exemplified in the blogosphere and in the growth of social networks centred around contents like MySpace[2], Wikipedia[3], YouTube[4] or Flickr[5].

In the current scenario centred around the concept of Web 2.0, the multimodal nature of the World Wide Web shows itself most clearly. Collaborative web sites for dissemination of video files, pictures, music and text have become widespread, and this in the multilingual setting which is typical of the web. In such a setting, a growing need arises for access and retrieval of multimodal information.

Despite the above mentioned multimodal nature of the web, the information, whatever its nature (image, video, music or text files), is indexed by text, that is to say, search engines rely on keywords to retrieve information regardless of its format.

This paper goes into the multimodal nature of the web, outlines new trends in information access and retrieval, and describes a system for multilingual retrieval of image files.

[1] <http://es.wikipedia.org/wiki/Web_2.0>.
[2] <http://www.myspace.com/>.
[3] <http://www.wikipedia.org/>.
[4] <http://www.youtube.com/>.
[5] <http://www.flickr.com/>.

Authors

Luis-Alfonso Ureña-López is a Senior Lecturer in the Department of Computer Science at Jaén University (Spain). He received a M.S. degree in Computer Science at Granada university in 1991, and a Ph. D. in Computer Science at Software Engineering Department of Granada University in 2000. His Ph. D. Thesis was awarded in the 2001 Awards of the Spanish Society for Natural Language Processing. From 1997 to 2004, he has been the head of the Computer Science Department at Jaén University. He is currently director and founder of the Intelligent Systems Information Access Research Group in the Computer Science Department. His current research includes Word Sense Disambiguation, Information Retrieval (multilingual and multimodal), Text Categorization and Management Systems of Natural Language Processing and Human Computer Interaction. He has led R&D projects under national programs and for private companies, having published more than 75 scientific papers. Referee of different program committees of international conferences and International Journals. Organization of different national and international conferences. <laurena@ujaen.es>.

Manuel-Carlos Díaz-Galiano is Assistant Professor in the Department of Computer Science at Jaén University (Spain). He received a M.S. degree in Computer Science at the University of Granada in 1990. His current research includes Information Retrieval, Multimodal Information Retrieval and Management Systems of Natural Language Processing and Human Computer Interaction. <mcdiaz@ujaen.es>.

Arturo Montejo-Ráez is Assistant Professor in the Department of Computer Science at Jaén University (Spain). He received a M.S. degree in Computer Science at the University of Jaén in 1999, and a Ph. D. in Computer Science at Software Engineering Department of Granada University in 2006. His current research includes Information Retrieval, Text Categorization and Management Systems of Natural Language Processing, and Spoken Language Dialogue Systems. <amontejo@ujaen.es>.

M. Teresa Martín-Valdivia is Assistant Professor in the Department of Computer Science at Jaén University (Spain). She received a M.S. degree in Computer Science at the University of Granada in 1992, and a Ph. D. in Computer Science at Software Engineering Department of Málaga University in 2004. Her research interests include the application of neural networks, text categorization, multilingual and multimodal information retrieval. <maite@ujaen.es>.

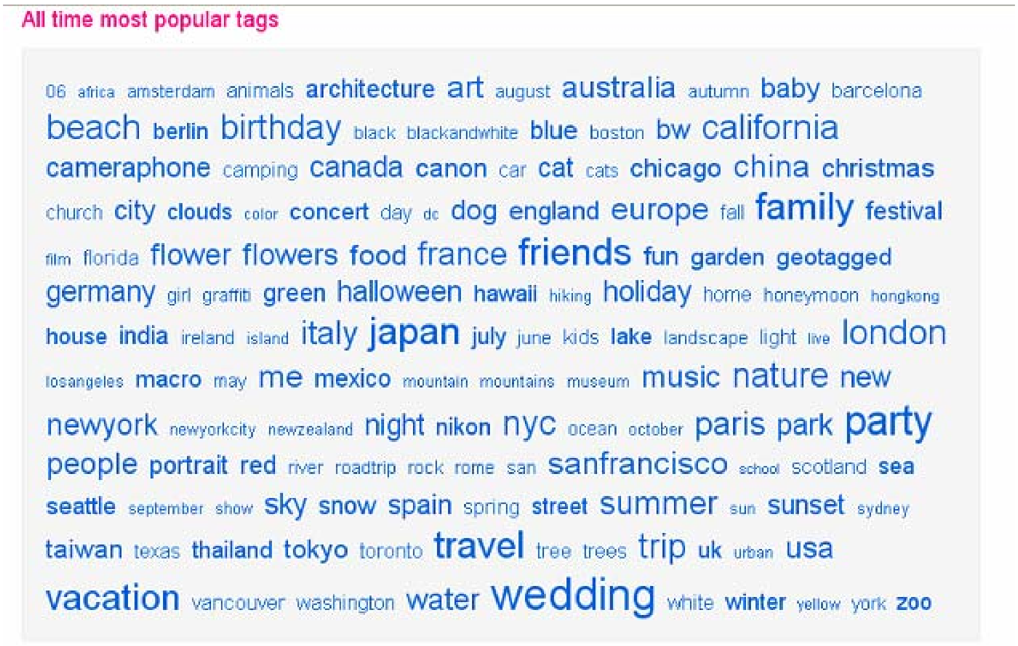


Figure 1: The Tags most widely Used by Flickr Users. Source: http://www.flickr.com/photos/tags/. Each picture stored is usually retrieved by this piece of meta-information.

2 Multimodality

From the point of view of the technologies involved in the information exchange over the now ubiquitous http protocol, the World Wide Web is in continuous evolution. The web grows as the number of people who are on the Internet grows, these people becoming new information providers and consumers. However, trying to understand how this growth takes place should be approached based on browser contents and on its publishers' needs, namely the web users, rather than on the underlying technologies.

Related results are retrieved and categorised according to their relevance. The ability to supply and consume this information is crucial in the Information Society, where cultural, political, economic and technological positioning is given precisely by that ability. However, information is useless if it cannot be synthesized, processed, adapted to our needs and shared for further consumption. The information should be presented in or linked to its context and kept up to date, dynamic, and be modifiable…, before *knowledge* can even be mentioned.

Web-technology-based tools have proved more successful whenever there has been a chance to generate knowledge. Wherever a piece of information could be related to the rest of available data, commented on, linked to users and then become the centre of user-created *communities*, more and more complex social networks have risen spontaneously. These networks share knowledge, not just information, because each contribution has its own life in the sense that it allows users to interact with it beyond mere retrieval.

Some systems have thus become popular very rapidly because they provide user-oriented rather than data-oriented support. Some of these systems double the number of their registered users month on month and have become web services which have spawned clone sites in virtually every country and language. This means communities where thousands and even millions of users share texts (blogs, forums, document repositories like Google Docs[6], collaborative encyclopedias like the Wikipedia, e-learning tools like Moodle[7]...), technical records (as in the Internet Movie Database[8]), video files (Google Video[9], YouTube o iFilm[10]), images (Flickr, Picassa[11]), audio files (a variety of *podcasts*), links (Del.icio.us[12], Blinklist[13] or Furl[14]), commercial sites (eBay[15] and Amazon.com[16]) and even georeferenced information (Google Earth, Virtual Earth). The varied nature and format give the web its multimodal character. Its contents are so varied that it is hard to tell whether the information which we may be interested in is an audio stream, a video or a text file. Future search engines will therefore be limited by their ability to cope with such polymorphic formats, and their quality and strengths will be measured thus too (see Figure 1).

But, was this not about knowledge? What became of all that? It has been mentioned above that it is the added value of users and their organizing social networks that make a

[6] <http://docs.google.com>.
[7] <http://moodle.org/>.
[8] <http://www.imdb.com>.
[9] <http://video.google.com>.
[10] <http://www.ifilm.com/>.
[11] <http://picasa.google.com/>.
[12] <http://del.icio.us>.
[13] <http://www.blinklist.com/>.
[14] <http://www.furl.net/>.
[15] <http://ebay.com>.
[16] <http://amazon.com>.

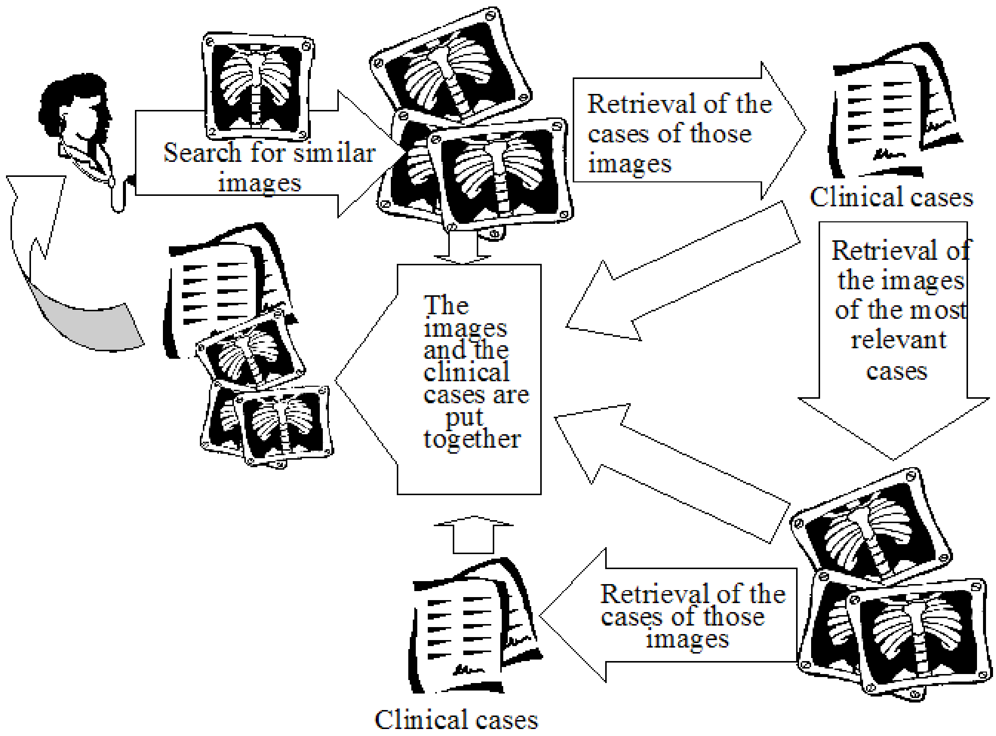


Figure 2: An Example of a Mixed Visual and Textual System (IR + CBIR).

difference. Take a video file hosted in YouTube. How can it be retrieved? While a direct link may be available in, for instance, an email from a friend, the fact is that this video file database must provide easy access to its contents. How this is achieved can be seen by looking at its website: users can see a selection of the most highly voted video files, browse the categories established by other users, jump from one video file to another according to their similarity and even use a search engine. But what makes this a rich collection? Simply the community-generated accompanying texts, the knowledge built around it. In this multimedia collection users can tag each contribution for easier identification of the contents as well as provide comments on the contents, rankings, etc. These tools are available in other collaborative environments such as blogs, shared links, movie databases and e-commerce. What makes these information collections valuable is the information added by users. This is the secret of Web 2.0's success [6]: the information becomes shared knowledge which can become easily widespread.

However, the major communication tool between users is *language*: it remains an essential communication tool in any object exchanged online. In fact, *metadata* consist always in small text fields. Meta-information is becoming more and more important (as in the access protocol to online libraries Open Archive Initiative, or the OWL ontologies). The Semantic Web [5] feeds on a formalization of metadata, so it can grow from the information generated by the systems inspired on Web 2.0 concepts. However, metadata is made up of rigid entities and their structure is pre-established and not user-friendly for those who may want to add occasional pieces of information along with perhaps a tag, a comment and a couple of links to other pieces of information on the web.

Therefore, it is still the ability to understand and use natural language that sets the limits of the processing of multimodal information. While it is true that "a picture is worth a thousand words", the web is still mainly textual and regardless of the information, whatever its format may be, it cannot be understood without a description, a remark or an associated piece of text. Language is still the wrapper for any object intended to become part of the global sphere of knowledge.

3 The Experience in Multilingual Image Retrieval

Countless systems rely on natural language for their queries against images. These systems are well-known and use many of the techniques of Natural Language Processing (NLP) and of Information Retrieval (IR), which are based on firmly grounded systems which use textual information alone.

At present, the interest in tools capable of retrieving images based on a given image is on the increase, as for example, person search on Flickr[17], automatic linguistic indexing and search of Pictures in ALIPR[18] or Google[19] Image. They are image retrieval systems based on contents, hence their acronym CBIR – *Content-Based Image Re-*

[17] <http://www.flickr.com/search/people/>.
[18] <http://www.alipr.com/>.
[19] <http://images.google.es/>.

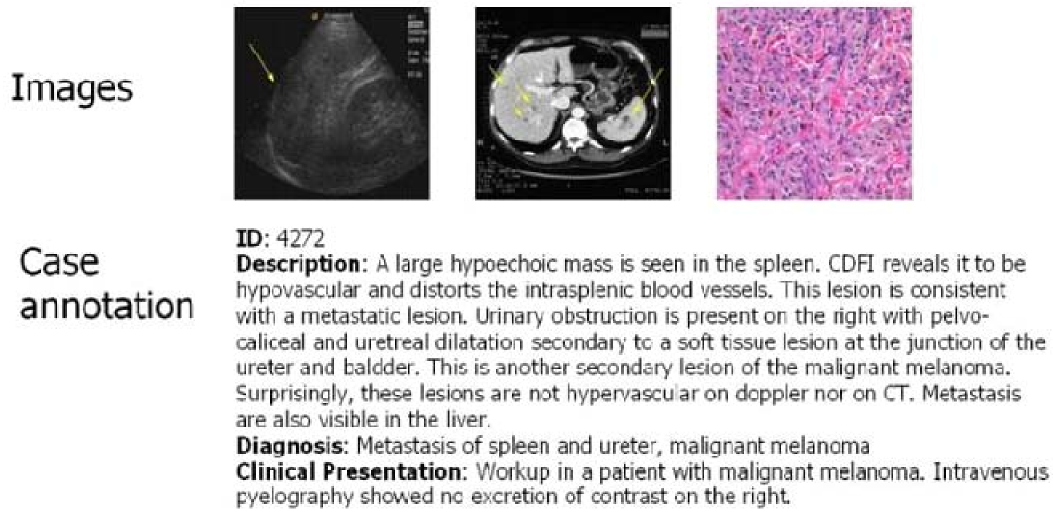


Figure 3: An Example of the Information of a Clinical Case <http://ir.ohsu.edu/image/2006protocol.html>.

trieval. A CBIR system indexes images according to their low-level visual features, like their texture, their colour or the geometrical forms contained in them. Some of the improved systems use techniques like image segmentation and rely on some image segments to solve the query, or use several segments as new queries and then collates the results. Despite the progress made in content extraction, CBIR systems are still hindered by considerable limitations.

More often than not, CBIR systems do not rely solely on visual information. Many information collections offer, beside images, some kind of textual information (free or semi-structured text) which has a semantic relationship to the image, for example, history archives, art collections or medical databases. Image retrieval systems can also give better results using the textual information associated with images.

3.1 Medical Tasks

Medical science often has cases which include images, text notes or reports written by specialists that are all contained in the patients' medical files. A traditional system of textual retrieval could benefit considerably from a CBIR system by retrieving cases similar to the ones which are most relevant as obtained through an image-based query.

Similarly, a CBIR system could give better results if it used all the textual information contained in a medical file, not just for retrieval of similar or identical images, but as a supplement to the information on the clinical case related to those images. This would help specialists gain access to all the information of the disease which those images portray (see Figure 2).

Our approach to this perspective relies on textual information for improved CBIR systems. Our experiments test various techniques for combined use of visual and textual information in a variety of clinical cases. However, the best experiments are often those which rely on the simplest ideas, like merely putting together images retrieved by various retrieval systems (visual and text).

Our techniques were tested in the past two CLEF[20] [1][2] campaigns, specifically in the subtask ImageCLEFmed[21],

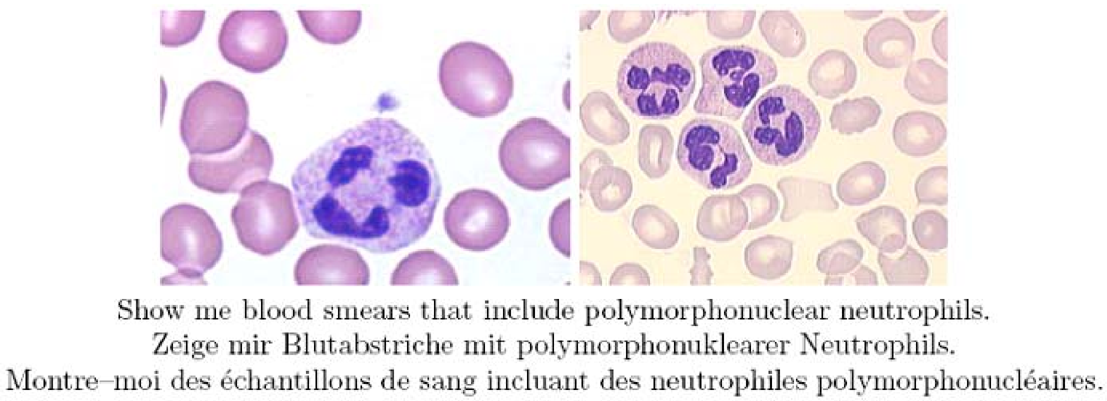
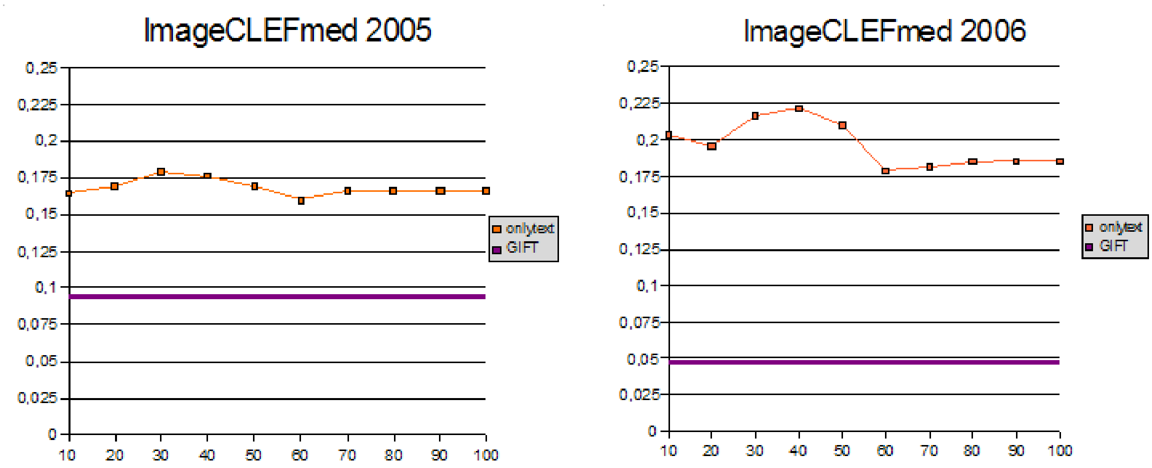
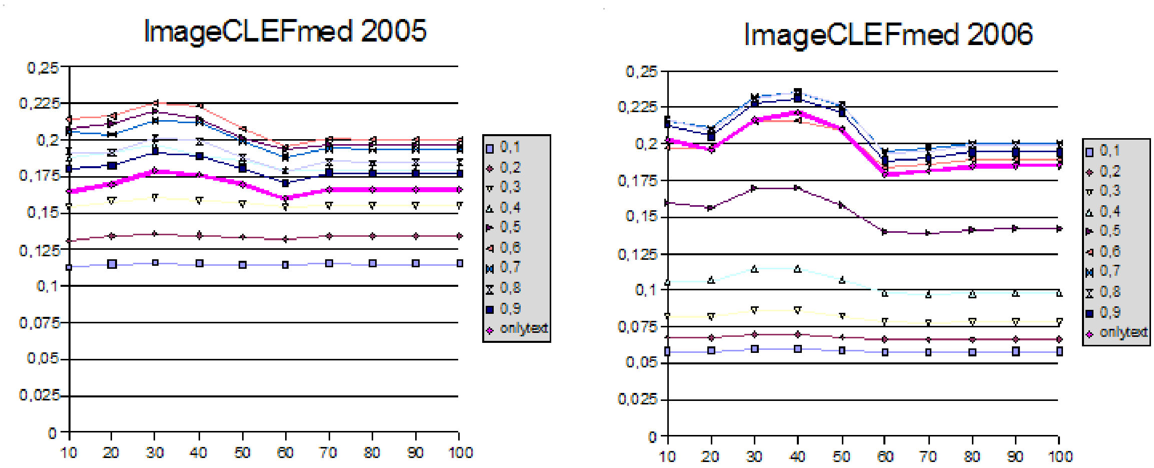


Figure 4: Example of Query in ImageCLEFmed 2006.



(a)



(b)

Figure 5: A Comparison of Text Retrieval and Image Retrieval, a) Visual (GIFT) versus Textual (onlytext), b) Using different Weightings.

for independent assessment of the system and one which relies on widespread resources.

The collection of clinical cases available in the CLEF [3] includes images of four different types of databases (MIR, PEIR, PathoPIC and Casimage). They contain in the range of 50,000 images along with their matching XML-annotated cases (see Figure 3).

Beside work collections, a CBIR tool called GIFT is also available. Our system uses that software for a preliminary retrieval of images for further refinement.

Each CLEF campaign makes available for participants a number of queries consisting of an image and a short text (in three languages: English, German and French), and a query. An example of such a query is shown in Figure 4.

3.2 The Resources

The body of clinical cases is very heterogeneous. The collections PEIR and PathoPIC contain medical images (scans, x-ray, CT, MRI, …) with notes for each image. Instead, the collections MIR and Casimage are arranged by clinical cases. Each case contains a number of medical images and notes on the disease depicted in the images.

In addition, not all cases are in the same language: although most of the contents are in English, some collections are in another language or are translations. Only 20% of the cases in the Casimage collection are in English. It was decided to translate the remainder from French into English. The PathoPIC collection contains cases either in German (the original versions) or English (translations from German).

3.3 The Experiments

The first problem for our system was having to use heterogeneous collections, so the collections were normalized as far as possible: the collections were first pre-processed for XML annotation, because the collection MIR was not tagged in full (even though the same structure was used throughout).

The contents of tags which did not supply any relevant information were then deleted for reduced noise. The text

[20] <http://www.clef-campaign.org/>.
[21] <http://ir.ohsu.edu/image/>.

fields of the files of each case which could supply most information based on the content and certain statistical measures were then selected. The efficiency of the deletion of certain clinical fields was tested by generating several collections using a given percentage of the labels with the best information. Therefore, collections were generated using a given percentage of the labels with the best information. The percentages used were: 10%, 20%, 30%… , 90%, 100%.

For each of the collections generated, image-text pairs were created by retrieving text information about the images from the cases in which they were contained. Simple text retrieval was thus possible, because each retrieved document matched only one image.

At this stage, the visual and text information are combined. Where queries consist of an image plus text, each element is used as queries in different systems: images are searched for using GIFT, and each image query returns a list of relevant images; texts are searched using LEMUR, and a list of relevant documents is returned too. The lists are lists of relevant documents, because each document has a matching image and vice versa.

Both lists are put together giving a certain weight to each. Different weights were given to each list in order to test how this influenced the results. The weights given to the text list were 10% (0,1), 20% (0,2), …, 90% (that is, 9 more types of tests).

3.4 The Assessment

The results obtained were assessed based on the (relevance) RELEVANT judgements provided by the organization of CLEF. The image and text results (retrieved with GIFT and LEMUR[22], respectively) were assessed separately so they could be used as base cases. This made it possible to identify which of the two systems yielded better results too. All text queries were submitted to the 10 text collections generated, each of them with different tag percentages. The mixed lists were also assessed with varying weights for the text list (0.1 - … - 0.9).

3.5 The Results

Very similar results were obtained for the query list of ImageCLEF 2005 and 2006. In both cases, text retrieval yielded better results than image retrieval (Figure 5a). Thus, the mixed list (textual + visual) produces better results when the contribution of the textual list has more weight than that of the visual list, being these results even better than those obtained by using just textual information (Figure 5b).

As to information retrieval-based tag selection, better results were obtained when fewer tags were used. The best results were obtained when 30% to 50% of the best tags were used. This can be noticed even when image and text information are together.

[22] <http://www.gnu.org/software/gift/>.

[23] Sistema de Recuperación de Información <http://www.lemurproject.org>.

4 Conclusions

Web content is growing exponentially. We are currently shifting from information to knowledge through the exchange and interaction of information by user communities. Knowledge is an entity which is dynamic and which is interactive, and it is managed by the tools enabled by Web 2.0. Thus, it does not matter which object is exchanged, (image, video, sounds), because it is associated text that turns it into global knowledge and gives it its place in the stream of ideas, data, opinions, scorings, profiles and countless other elements which make up these large collaborative knowledge databases.

The challenge is now whether researchers can make knowledge management and generation easier based on that information. With the presence of associated texts, NLP techniques can retrieve from the blogosphere the underlying social network or solve a query of clinical images based on a given case. Research on information retrieval will have to manage the overwhelming amount of information with faster, more efficient and more specialized tools, but always with the guarantee that, deep down, they can always rely on words. Language is an inherently human feature, and its use will remain the cornerstone of knowledge. Our ability to process it will set the limits to our information technology and to knowledge.

Acknowledgements

We would like to thank the organizing committee of CLEF, specifically Carol Peters, for access to the CLEF resources and subsequent referencing. This work has been supported by the Spanish Government with grant TIN2006-15265-C06 (TIMOM).

References

[1] M. T. Martín-Valdivia, M. A. García-Cumbreras, M. C. Díaz-Galiano, L. A. Ureña-López, and A. Montejo-Ráez; The University of Jaén at ImageCLEF 2005: Adhoc and Medical Tasks. LNCS 4022, pp 612-621, 2006. Springer Berlin / Heidelberg.

[2] M. Teresa Martín-Valdivia, Miguel A. García Cumbreras, Manuel C. Díaz-Galiano, L. Alfonso Ureña López, and Arturo Montejo-Ráez. SINAI at ImageCLEF 2006. Working Notes CLEF 2006 (Cross Language Evaluation Forum).

[3] Henning Müller, Thomas Deselaers, Thomas Lehmann, Paul Clough, and William Hersh. Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks. Evaluation of Multilingual and Multi-modal Information Retrieval — Seventh Workshop of the Cross-Language Evaluation Forum, CLEF 2006. LNCS 2006, Alicante, Spain.

[4] Tim Berners-Lee and Mark Fischetti (1999). Weaving the Web: Origins and Future of the World Wide Web. Britain: Orion Business. ISBN 0-7528-2090-7.

[5] T. Berners-Lee. Semantic web road map. Technical report, W3C Design Issues, 1998.

[6] Tim O'Reilly. *What is Web 2.0*. Oreillynet.com, 2005. <http://www.oreilly.com/go/web2>.

