Comparing the quality of accessing the medical literature using content–based visual and textual information retrieval

Henning Müller[ad], Jayashree Kalpathy–Cramer[b], Charles E. Kahn Jr.[c], William Hersh[b]

[a]Service of Medical Informatics, University Hospitals & University of Geneva, 24, rue Micheli–du–Crest, CH–1211 Geneva 14, Switzerland;
[b]Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA;
[c]Department of Radiology, Medical College of Wisconsin, 9200 W. Wisconsin Avenue, Milwaukee, WI 53226-3522, USA
[d]University of Applied Sciences Western Switzerland (HES SO), TecnoArk 3, 3960 Sierre, Switzerland

ABSTRACT

Images play an important role in medical diagnosis and treatment planning. When looking at tools for diagnostic aid it becomes clear that textual retrieval methods are much more often used than visual ones to supply additional information for a case under observation. Content–based visual image retrieval has been proposed many times to aid diagnosis in specific domains.[1] Access to the scientific literature was also already proposed[2] but without an actual implementation and evaluation of the results.

The ImageCLEFmed* benchmark on medical information retrieval makes available each year a collection of images with annotations, potential informations needs of clinicians, and a gold standard for the evaluation of the quality of research algorithms using visual and textual retrieval. In general, around 35 research groups inscribe for the task every year and 15 finally submit results to compare their techniques.

In this paper we describe access to a dataset of images from journals of the Radiological Society of North America (RSNA) in the context of ImageCLEFmed 2008. The images were made available in collaboration with the ARRS GoldMiner®† search engine, where all images can be queried via a web interface.[3] Articles are available in full text on the web as well. Figure captions and the part of the caption referring to a particular image of the figure were also made available to participants.

In total, 37 research groups from 25 countries registered for the task and 15 submitted results to all 30 query topics. A large variety of techniques was used from purely visual retrieval to purely textual retrieval and mixed techniques. Interaction strategies such as relevance feedback and automatic query expansion were also tested. Query topics were supplied in three languages (English, French, German) so multilingual retrieval from a collection in English was evaluated. The results show that visual retrieval and particularly when combined with textual retrieval can supply access to the relevant medical literature to fulfill the clinician's information needs. Textual retrieval alone has much better results than visual retrieval alone. For future competitions it is foreseen to move from the notion of relevant images to fulfill an information need rather towards the notion of relevant cases.

Keywords: scientific literature, medical image retrieval, evaluation, information retrieval

Further author information: (Send correspondence to Henning Müller, Email: henning.mueller@sim.hcuge.ch

*http://www.imageclef.org/

†http://goldminer.arrs.org/

1. DESCRIPTION OF PURPOSE

Goal of the ImageCLEF benchmark (part of CLEF[‡] — Cross Language Evaluation Forum) is to compare multilingual information retrieval systems. A subtask of ImageCLEF that is described in this abstract deals with the retrieval of medical images from a collection of images taken from the radiological journals Radiology and Radiographics. 30 query topics were developed containing a text of the information need in three languages and with two example images.

Main goal of the ImageCLEF benchnmark is to make available for visual and textual information retrieval researchers a platform to compare their algorithms with those of other researchers to identify promising techniques by comparing them on the same data and tasks.

2. METHODS

Thirty query topics were developed based on surveys among medical professionals, and the analysis of query logs from the Medline literature search engine and the hon (health on the net) medical media search engine. These query topics correspond thus to realistic information needs of clinicians. Query topics contain images, and the textual information need in three languages (English, French, German).

The dataset contains 67'000 images made available from the GoldMiner image search engine. These images are from the journals Radiology and Radiographics. Together with the images the part of the figure caption concerning a particular image, the full caption for all images of a figure, and a link to the full text article in English were made available.

Based on the submitted runs of the research groups, pools of images were judged for relevance or irrelevance by medical doctors that were also students at the OHSU medical informatics program. The submitted runs were then compared and a workshop to discuss the results is planned for September 2008 in Aarhus, Denmark.

3. RESULTS

30 realistic medical query topics were created in ImageCLEFmed 2008, and these were distributed among registered participants together with a database of 67'000 images taken from the medical literature. 37 research groups registered for ImageCLEFmed 2008, and 15 groups finally submitted results using a large variety of techniques. Medical doctors judged the images of the submitted runs for relevance, and thus a gold standard for evaluation was created. In total, 130 runs of these 15 research groups were compared. Comparison of the results allowed to identify the best–performing visual techniques, textual techniques and also methods for combining visual and textual cues for retrieval. The organization of the workshop in connection with the retrieval challenge gives researchers a forum to discuss approaches and compare techniques for optimizing their performance.

4. NEW OR BREAKTHROUGH RESULTS

Visual and textual image retrieval alone and combined were compared on 30 query topics and for 15 research systems with a database of 67'000 images from the journals *Radiology* and *RadioGraphics*. This permits an objective overview on the techniques that can help clinicians to fulfill their information needs from the medical literature. Promising techniques can thus be identified and be followed to improve their retrieval quality.

This is the first study to examine and evaluate content–based visual data access to the medical literature. Comparing 15 research systems on this task allows a good overview of currently obtainable quality.

[‡] `http://www.clef-campaign.org`

5. CONCLUSIONS

Benchmarks to evaluate and compare research techniques are extremely precious for the international research community as they allow for many research groups not linked to e medical institution to have access to large and realistic data sets without the need to recreate these data set from scratch each time.[4] These commonly used datasets also make research techniques comparable and allow groups to improve their research outcomes by following the most promising techniques and approaches. ImageCLEFmed has made available databases, search tasks, and ground truth for five years now, and over 100 international research groups have inscribed for these tasks underlining the need for such efforts their usefulness.

6. SUBMITTED ELSEWHERE

Other publications on ImageCLEF exist[4–6] but this is the first article using the current database of images from the medical literature being used in ImageCLEF. There is thus no other publication very similar to this one.

7. ACKNOWLEDGEMENTS

This work was partially supported by the University of Applied Sciences Western Switzerland (HES SO) in the context of the BeMeVIS project. The relevance judgments were supported by the National Science Foundations and Google Inc. The images for the 2008 ImageCLEFmed challenge were contributed by the Radiological Society of North America (RSNA).

REFERENCES

1. H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content–based image retrieval systems in medicine – clinical benefits and future directions," *International Journal of Medical Informatics* **73**, pp. 1–23, 2004.
2. T. M. Deserno, S. Antani, and R. Long, "Exploring access to scientific literature using content-based image retrieval," in *SPIE Medical Imaging*, **6516**, (San Diego, CA, USA), February 2007.
3. C. E. Kahn Jr. and C. Thao, "Goldminer: A radiology image search engine," *American Journal of Roentgenology* **188**, pp. 1475–1478, 2008.
4. W. Hersh, H. Müller, J. Jensen, J. Yang, P. Gorman, and P. Ruch, "Advancing biomedical image retrieval: Development and analysis of a test collection," *Journal of the American Medical Informatics Association* **September/October**, pp. 488–496, 2006.
5. H. Müller, T. Deselaers, E. Kim, J. Kalpathy-Cramer, T. M. Deserno, P. Clough, and W. Hersh, "Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks," in *CLEF 2007 Proceedings*, *Lecture Notes in Computer Science (LNCS)* **5152**, Springer, (Budapest, Hungary), 2008.
6. T. Deselaers, H. Müller, P. Clough, H. Ney, and T. M. Lehmann, "The CLEF 2005 automatic medical image annotation task," *International Journal in Computer Vision* **74**(1), pp. 51–58, 2007.