Accessing the Medical Literature with Content-based Visual Retrieval and Text Retrieval Techniques

PARTICIPANTS
Dimitrios Markonis
Adrien Depeursinge
Ivan Eggel
Antonio Foncubierta-Rodriguez
Henning Mueller PhD

Background

Radiology is strongly connected with image use and search. However, the management and reuse of the overwhelming amount of medical image data produced by hospitals is not yet satisfactory. Search by using patient IDs or keywords does not allow exploiting the full richness of large databases of images with attached diagnoses due to scarce and often erroneous annotations available, for example in DICOM headers. This paper presents a medical literature search system that combines textual and visual content-based retrieval with relevance feedback aiming at speeding up the iterative search process and improving the retrieval results.

Evaluation

The proposed system is based on two main components: Lucene for the text retrieval part and the GNU Image Finding Tool (GIFT) for content-based visual information retrieval. The user front-end is a simple web interface based on Java Server Faces (JSF) (see Figure). The database of the medical image retrieval task in ImageCLEF2011 was indexed and contains approximately 55,000 articles and 228,000 images of the open access biomedical literature. The initial query can be either a keyword or an image. The result consists of a list of ranked articles along with thumbnails of images contained in each paper.

Discussion

Previous studies have shown that content-based image retrieval achieves high early precision while text retrieval has much better recall. To benefit from the two approaches, this work uses relevance feedback to combine content-based queries to the initial textual or visual query. The simple web interface enables fast and easy access to large amounts of images in the medical literature. The results format showing image thumbnails allows the user to quickly check for relevance of the retrieved articles and refine the query using content-based queries from returned images.

CONCLUSION

The task of medical document and image search is highly time-consuming for radiologists and often leads to unsatisfactory results. In this work, a medical literature retrieval system is proposed, combining textual and visual information querying to enhance the search process and provide fast and straightforward access to the increasing amount of produced medical information.