MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-1273

A Web Service for enabling Medical Image Retrieval Integrated into a Social Medical Image Sharing Platform

Marko Niinimäki[a], Xin Zhou[a], Enrique de la Vega[a], Miguel Cabrer[c], Henning Müller[ab]

[a] *Service of Medical Informatics, Geneva University Hospitals and University of Geneva, Switzerland*
[b] *Business Information Systems, University of Applied Sciences Western Switzerland (HES–SO), Sierre, CH*
[c] *MEDTING Medical Exchange, Dublin, Ireland*

Abstract

Content–based visual image access is in the process from a research domain towards real applications. So far, most image retrieval applications have been in one specialized domain such as lung CTs as diagnosis aid or for classification of general images based on anatomic region, modality, and view. This article describes the use of a content–based image retrieval system in connection with the medical image sharing platform MEDTING, so a data set with a very large variety. Similarity retrieval is possible for all cases of the social image sharing platform, so cases can be linked by either visual similarity or similarity in keywords. The visual retrieval search is based on the GIFT (GNU Image Finding Tool). The technology for updating the index with new images added by users employs RSS (Really Simple Syndication) feeds. The ARC (Advanced Resource Connector) middleware is used for the implementation of a web service for similarity retrieval, simplifying the integration of this service. Novelty of this article is the application/integration and image updating strategy. Retrieval methods themselves employ existing techniques that are all open source and can easily be reproduced.

Keywords:
Image retrieval, Medical social web, Visual information retrieval.

Introduction

Images are produced in enormous quantities in all modern hospitals, and they are an essential part of diagnosis and treatment planning [1]. They are also increasingly accessible to non–imaging specialists in hospitals directly through the electronic patient record creating a need for a diagnosis aid based on the images. To better manage the information and extract knowledge from existing cases, new tools are required. Content–based image retrieval (CBIR, [2]) has been proposed in the medical domain several times [3,4]. Still, real use of medical CBIR has been limited to a few examples [5] and most often purely academic tasks such as image classification have been performed [6]. In general, CBIR extracts visual features from the images and then allows for searching visually similar images to one or several examples given, sometimes relevant and irrelevant examples can be given. One of the current problems of image retrieval is the integration of research tools into existing applications such as viewing stations or directly into the patient record. In such an integrated case image retrieval can give access to anonymous and annotated cases from the past (so from the patient record or the literature) to solve a current case. There are of course several hurdles regarding secondary data use to overcome such as legal and ethical questions [7].

The GNU Image Finding Tool used as a visual retrieval engine has an interface for querying and receiving results. The interface language of GIFT is called MRML[1] (Multimedia Retrieval Markup Language) and allows for connecting to a server, choosing databases, parameters for algorithms, and receiving similarity results including identifiers and scores for images. The current implementation has several drawbacks:

• GIFT uses socket–based communication, typically on a port blocked by firewalls.

• MRML is based on XML but its handling is cumbersome and does not follow current standards as it was developed over ten years ago.

In this paper, we present a solution for adding a web service for similarity–based image retrieval with example images that is easy to integrate. The solution follows standards and communicates through port 80, often open in networks for WWW access. A medical Web 2.0 application for sharing case reports (often from radiology) including images integrated this web service to allow for finding similar cases not only based on key words but also on similar images regarding visual similarity. The web portal of MEDTING[2] is commonly used by clinicians and particularly radiologists. For our application, two tasks were implemented:

(1) cumulative indexing of newly added images, and
(2) a web service for finding similar images based on visual similarity.

Cumulative indexing is needed since new images are added daily and need to be integrated into the collection in a simple way. For updating our database MEDTING generates an RSS (Really Simple Syndication) feed containing image URLs (Uniform Resource Locators). A small program reads the feed

[1] http://www.mrml.org/
[2] http://www.medting.com/

daily and indexes any new images. The motivation for a web service is the provision of a simple but flexible interface with WSDL (Web Service Description Language) that is absolutely standard and can be integrated easily in all programming languages.

Materials and Methods

Tools reused

All tools reused for the described implementation are open source and available free of charge, so results can be reproduced easily. The GIFT[3] [8] is an open source tool for visual retrieval in large photo collections. Color plays an important role but GIFT has also been applied for medical image retrieval from collections containing grey–scale images. GIFT employs four groups of visual features for retrieval. Texture features based on Gabor filter responses exist as local features of small blocks and as global features in the form of a global histogram. Color or grey level features in the HSV (Hue, Saturation, Value) color space equally exist as local blocks at various scales and as a global histogram. GIFT uses many techniques from text retrieval such as frequency–based feature weightings and relevance feedback techniques. This creates a fast retrieval of under 0.5 seconds for databases of over 100'000 images on standard desktop computers. MRML is an XML–based query language that allows connecting to a GIFT server.

MEDTING is a social Web 2.0 service allowing clinicians and lecturers to share medical cases including images. Cases can easily be added and annotated via an ontology. Similar cases are proposed to the user when browsing and also comments on particular cases made by other users. The web side allows for a fast navigation including simple image viewing. Started only in 2008, the portal has reached in spring 2009 over 1'800 registered users. Currently, 2'400 clinical cases are stored on the portal with over 20'000 images and over 10'000 accesses per week. Such a scenario requires a stable and fully automatic solution to limit integration work, leaving mainly the possibility to place a button "Show me similar images" next to the images in the interface.

KnowARC[4] (Knowhow sharing with the Advanced Resource Connector) is a project developing a Grid middleware used for computationally expensive tasks in the physics domain but also in medical imaging. One goal of the project is to integrate existing medical applications into clinical real–world applications showing the potential of Grid technologies to make existing research tools faster and more effective. The Grid technologies are mainly used for the offline feature extraction in our case and not for the online querying that is already fast. The ARC [9] middleware of the project offers several tools that are useful, for example to ease the creation of web services.

Database used

The database used in the work described in this article includes all images currently stored on the MEDTING web pages. By summer 2009, well over 20'000 images from more than 2'000 cases are indexed and thus taken into account for similarity retrieval. The images are extremely varied and include almost all medical imaging modalities, drawings, and photographs. The number of images is continually rising. There is no ground truth available to evaluate the performance of retrieval on this database. All employed tools have been evaluated on standard databases for their quality, though.

Results

Main result of this article is not a new retrieval technique but the integration of an existing system (GIFT) into a new application and the creation of a framework for this integration. The same techniques have been used for integrating the same image retrieval application into the viewing stations of the Geneva University Hospital in the past. More on the retrieval quality using GIFT on a standard database can be found on the ImageCLEF[5] web pages.

To limit the amount of integration work we decided to use web services. Web services [10] provide a language independent interface. A useful feature of web services is the use of the WSDL (Web Service Description Language) that simplifies client development. As a web service container (a server that provides the services), we used a component of ARC, a recent development. The web service 'FMEDTING' is invoked by MEDTING when the user views an image of interest and clicks the "Find similar" button. The button calls a client module. fname is the name of the source image file, for instance v_tmp_22042009_14121159.jpg:

```
\$client=newSoapClient(''FMEDTING.wsdl'')
\$res=\$client->image(array('imagefile'=>\$fname));
```

The functionality of the web service can be described as follows:

- The service is invoked with a file name (parameter "imagefile"). Optional parameters indicate the number of results and a similarity threshold.
- The service initiates a connection with GIFT. The IP address of the server and the port number are configured in the configuration file.
- The service sends a query to GIFT, receives result and forwards them to the caller.

[3] http://www.gnu.org/software/gift/
[4] http://www.knowarc.eu/
[5] http://www.imageclef.org/

Figure 1 - Screenshot of MEDTING including the button to search for similar images.

Since the web service itself is straightforward, it adds no overhead compared to a situation where GIFT is called directly via socket–based communication. The client's overhead consists of generating the service call from the WSDL file and interpreting the result. On a modern desktop computer, these add 0.2 seconds. Calling the service with a client and receiving the image URLs of the 10 most similar images takes 1.5–2.5 seconds. Image identifiers in GIFT are URLs, and thus the integration is easy using the URLs of images on the MEDTING web site. A screenshot of the functionality integrated into MEDTING can be seen in Figure 1 with the "Show me similar images" button. Figure 2 shows the result of similar images shown to the user and allowing to navigate to other cases with visually similar images.

After an integration of the image retrieval functionality into MEDTING it became clear that besides the retrieval a mechanism was needed to keep the database up to date every day. RSS (Really Simple Syndication) is an XML–based format that allows developers to describe and syndicate content [11]. The syndicated content, called feed, can consist the content itself, or its metadata. MEDTING started to provide an RSS feed containing the images added to its web side towards the end of 2008. The RSS feed contains URLs pointing to two versions (full size and thumbnail) of the images:

```
<rss version="2.0"
xmlns:media="http://search.yahoo.com/mrss"
xmlns:atom=''http://www.w3.org/2005/Atom''>
<channel>
<atom:link rel="next"
href=''http://medting.com/rss/resources.php?
page=2&amp;orderby=lastaddedresources''/>
<item>
<title>v_tmp_22042009_14121159.jpg</title
>
<link>http://medting.com/atlas/view.php?id=2
8428</link>
<media:thumbnail url=''http://media.medting.
com/28/28428_3VGn5Ih7IcoFYo_t.jpg''/>
<media:content url=''http://media.medting.
com/28/28428_3VGn5Ih7IcoFYo_e.jpg''/>
</item>
```

The RSS feed is read daily. If new images are discovered, they are added to the collection to always keep the files up to date. An average of 30 images have been added daily since February 2009. An overview of the entire system is shown in Figure 3.

Conclusion

The work described in this article details the integration challenges when using a research prototype for medical visual information retrieval in the context of a social medical image sharing site. The use of a web service for this simplifies the inclusion of finding similar images. The application is located at a University server and only the similarity search service is provided to the MEDTING web site. An important part of the integration was the automatic addition of new images, and using RSS provides a lightweight solution. For the integration of research prototypes into real applications it is particularly important to provide simple interfaces low in maintenance. This is surely more important than pure retrieval quality. For applications such as a permanently accessible social web site it is also important that the application is available all the time. We sometimes had maintenance stops or changes in the University server room resulting in a downtime and causing errors. In a next step we plan to analyze the usage logs of the system, finding out in which scenarios users search for visually similar images and which key word searches were performed with the system ahead of visual search. Having data of real, routine use can help understand situations in which users prefer to search visually and where key words are preferred. The created prototype allows obtaining direct feedback on visual search, important for a domain going from research to routine applications.

Figure 2 – Visually similar images to an example are shown under the image.

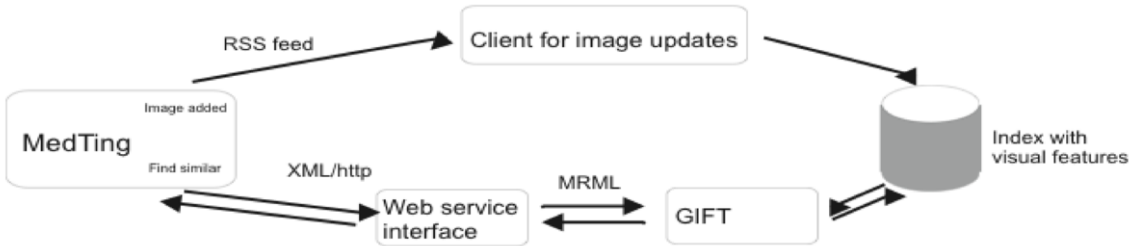

Figure 3-A global view of the system including the update mechanism using RSS feeds and the retrieval functionality integrated with a web service.

Acknowledgements

This work was partly supported by the FNS (200020–118638/1) and the EU FP6 KnowARC project (IST 032691).

References

[1] Haux R. Hospital information systems — past, present, future. International Journal of Medical Informatics, 75:268–281, 2005.

[2] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. Content–based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 No 12:1349–1380, 2000.

[3] Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content–based image retrieval systems in medicine – clinical benefits and future directions. International Journal of Medical Informatics, 73:1–23, 2004.

[4] Lowe H, Antipov I, Hersh W, Arnott Smith C. Towards knowledge–based retrieval of medical images. The role of semantic indexing, image content representation and knowledge–based retrieval. In Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA), pages 882–886, Nashville, TN, USA, October 1998.

[5] Aisen AM, Broderick LS, Winer-Muram H, Brodley CE, Kak AC, Pavlopoulou C, Dy J, Shyu CR, Marchiori A. Automated storage and retrieval of thin–section CT images to assist diagnosis: System description and preliminary assessment. Radiology, 228:265–270, 2003.

[6] Lehmann TM, Güld MO, Deselaers T, Keysers D, Schubert H, Spitzer K, Ney H, Wein BB. Automatic categorization of medical images for content–based retrieval and data mining. Computerized Medical Imaging and Graphics, 29(2-3):143–155, 2005.

[7] Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Detmer DE. Toward a national framework for the secondary use of health data: An american medical informatics association white paper. Methods of Information in Medicine, 14:1–9, 2007.

[8] Squire DM, Müller W, Müller H, Pun T. Content–based query of image databases: inspirations from text retrieval. Pattern Recognition Letters (Selected Papers from The 11th Scandinavian Conference on Image Analysis SCIA '99), 21(13–14):1193–1198, 2000. B.K. Ersboll, P. Johansen, Eds.

[9] Ellert M, Grønager M, Konstantinov A, Kónya B, Lindemann J, Livenson I, Langgaard Nielsen J, Niinimäki M, Smirnova O, Wäänänen A. Advanced resource connector middleware for lightweight computational Grids. Future Generation computer systems, 23(2):219–240, 2007.

[10] Curbera F, Duftler M, Khalaf R, Nagy W, Mukhi N, Weerawarana S. Unraveling the web services web: An introduction to soap, wsdl, and uddi. IEEE Internet Computing, March-April 2002.

[11] Hammersley B. Content Syndication with RSS. O'Reilly, 2003.

Address for correspondence

Prof. Dr. Henning Müller
Business Information Systems
University of Applied Sciences Western Switzerland
TechnoArk 3
3960 Sierre, Switzerland
tel ++41 27 606 9036
fax ++41 27 606 9000
henning.mueller@hevs.ch