Medical Image Retrieval: ISSR at CLEF 2010

Ragia Ibrahim, Waleed Arafa

Department of Compute & Information Sciences
Institute of Statistical Studies and Research (ISSR), Cairo University, Egypt
{ribrahim, waleed_arafa}@issr.cu.edu.eg

Abstract. This is the second participation of Institute of Statistical Studies and Research (ISSR) group in CLEF 2010-Medical image retrieval track. This paper describes our experiments in monolingual and multilingual tasks. First, we test Paragraph Extraction (PE) and Sentence Selection (SS) approaches on the classical medical retrieval task (Ad-hoc), as well as on Case-based retrieval. Second, we compare between three Cross Language Information Retrieval (CLIR) methods. These methods are Machine Translation (MT), dictionary translation as well as translating via thesauri. For indexing and retrieval, we used the Lemur toolkit. Regarding ad-hoc retrieval task best results obtained when image caption and title used only, and for case-based task, there is no significant difference between adding extra text to the article and using its title and its image captions. For multilingual task, there is no significant difference between the three methods.

Keywords: Information retrieval, cross-language information retrieval, textual retrieval, medical retrieval, text extraction, linguistic processing.

1 Introduction

Recently most doctors' computers contain medical images and descriptions for each image; also medical web sites provide mixed articles that discuss patients cases integrated with images for those cases. Medical images play a role in diagnosis process, since it based on comparing current case with previous similar cases, thus there is growing need for searching and retrieving medical images tool [1].

Medical images are typically retrieved by using; Content Based Image Retrieval (CBIR) that depends on selected image features (visual features), Context based retrieval that depends on image associated text (textual features), or by using mixture of textual and visual features. Also medical image associated text can be written in more than one language (multilingual), and the language used to express the textual queries should not be related to a specific language. The medical image system has to support finding images even though associated text written in one natural language and queries expressed in other languages [2].

Regarding monolingual retrieval our experiment focus on comparing between different selected textual features; three collections are created: the first collection used image caption and title only, the second collection used caption title and added paragraph, finally the third collection used caption title and selected sentence.

According to imageCLEFmed working notes several experiments concentrate on crossing language boundary as in [4, 5, 6 and 7]. This year we examine several cross-language information retrieval methods. These methods include using medical dictionary, readymade machine translation system and thesauri.

This paper is organized as follows: section 2 system description, section 3 ad-hoc retrieval experiments and discussion, section 4 case- based retrieval experiments and discussion, and finally the conclusion.

2 System Description

Institute of Statistical Studies and Research (ISSR) group IR system consists of the following components: indexing and retrieval, paragraph extraction, sentence selection and query translation component. For indexing and retrieval part, Lemur toolkit [8] is used.

Lemur was originally developed as part of the Lemur project, a collaboration between Language Technology Institute (LTI) at Carnegie Mellon University and Center of Intelligent Information Retrieval (CIIR) at the University of Massachusetts, Amherst. Lemur is an open source toolkit. It was written in C++ for Unix platforms, but also runs on Windows.

Sentence selection component aims to select the most relevant sentence from the added paragraph(s), and add it to the image file. To segment the paragraph into set of sentences we used a sentence segmentation tool, it is a natural language processing tool developed using Perl. From Cognitive Computation Group at the Department of Computer Science, University of Illinois at Urbana-Champaign. Sentence segmentation tool reads plain text file and rewrites it with one sentence per line [10]. Paragraph extraction and sentence selection methods are used to create textual collection for each image.

Regarding query translation, three Cross Language Information Retrieval (CLIR) techniques were experimented:

- Firstly, Machine Translation (MT) system is experimented. For this experiment, Google's free online language translation system is used. Thus, it is based on statistical machine translation. It supplies its translation system with billions of words of text, parallel text containing examples of human translations among the languages and monolingual text in the target language [12].
- Secondly, domain specific medical dictionary, for this experiment Reverso is used. It is a specialized online dictionary (French –English). It is a collaborative effort

between linguists, lexicographers, translators and specialists in the medical domain[13].

- Lastly, we experimented finding meaning using Unified Medical Language System (UMLS) Metathesaurus to translate French queries into English.

Unified Medical Language System (UMLS) is a repository of medical terms and their relationships, developed and maintained by National Library of Medicine (NLM). The NLM initiated UMLS in 1989, to fill the gap and to connect the individual vocabularies among each other to make an almost complete medical knowledge. UMLS Knowledge Sources consists of the Metathesaurus, the semantic network and the SPECIALIST Lexicon & Lexical Tools [11]. For our experiment, we used UMLS Metathesaurus. In CLIR experiment, the 2009AA release of the UMLS Knowledge Sources have been used , it contains more than 2.1 million concepts and 8 million unique concept names from over 140 source vocabularies, we specifically used English and French MeSH. The UMLS holds 295843 English MeSH Strings and 87000 MeSH French Stings belong to UMLS Metathesaurus.

2.1 Textual Collections Creation

Last year several retrieval models were tested. Best result obtained by using Indri indexing and Okapi language model. In addition a slight improvement is accomplished when updated stopword list is used, thus it used in all performed retrieval experiments for this year. We have also used our own developed Paragraph Extraction algorithm to add extra annotation to image before indexing [9]. From Mean Average Precision (MAP) point of view results, using only image caption and title performed better than using caption title and added paragraph(s). On the other hand, recall was increased. This year we constructed a new component that selects one sentence form extracted paragraph and add it to image file. The sentence selection method aims to select subset of previous extracted paragraph(s) sentences. Our aim is to add extra text related to the image and eliminate number of noisy terms, this done as follows:

- First, PubMed journal articles are downloaded, since distributed XML file includes article URL for each image. Then paragraph(s) relevant to each image is extracted.
- Second, each paragraph is segmented into sentences, using sentence boundary Perl script.
- Then similarity between each sentence included in the paragraph and image caption measured, using cosine similarity and TF/IDF weighting formula.
- After that, sentences ranked according to its similarity from image caption.
- Finally, the first ranked sentence is added to the caption and title of each image.

To compare between the use of textual features in ad-hoc retrieval task, we create three textual collections for each image.

CT collection: consists of each image caption and medical article title.
CTP collection: contains each image caption title and added paragraph(s).
CTS collection: contains caption title and selected sentence.

In order to participate in case-based retrieval task, the same three created collections were used but with grouping textual features. All images textual files that belong to the same PubMed journal article were concatenated in one file taking into account not to repeat article title or added textual segments. According to ImageCLEF 2010 website [14] the case-based retrieval task is different from the ad-hoc task, the unit of retrieval in case-based task is a case, not an image, therefore, PubMed journal article shall be retrieved. Thus, article URL is the unit of submission.

2.2 Query Translation

Indeed translation is required for CLIR; either translate documents to query language or translate query into document language. Clough concluded that query translation performance is better than document translation. However, results vary across languages and topics [14]. In represented experiments query is translated by using three methods. First, online Google MT system is used, second Reverso online medical French-English Dictionary is employed. In [18] general machine-readable dictionary is used to to translate query terms into document language. However, because of the lack of machine readable domain specific medical dictionary, Reverso is used and the translation to English is done manually; by extracting non-stop-word terms and translate them. Finally, query translation via UMLS Metathesaurus. Thesauri method depends on forming a variable length window, and then tries to translate the window or sub phrase. The variable length window is formed as follows:

- The entire query is considered as a window, the window length is the number of query terms. Then, translation technique is applied.
- If no translation found, the window length is decreased by one to create a new window. The new window starts at the first query term then it moves (shifts) right one term at a time until query end is reached. This happened until a translation is found. The shifting process is accomplished to construct sub phrases, then the system try to translate them. For instance, if we have a query of five terms {term 1, ... term 5} and translation failed for the five terms; new phrase will be formed, its size will be four and it will move on, this will give us two new windows {term 1…term 4} and {term 2, ..term 5}. If translation method failed again window size will be decreased into three, this will give us {(term1, ..term3), (term2, ..term4), (term3, ..term5)} and so on. The union of all translations window results will be the final translation.
- The process continues if there is any part or phrase not translated and all of its terms are not in stopword list, if window size reached size of two or one and all window terms in stopword list they are ignored.

In UMLS Metathesaurus, the entire concept structure appears in a single rich format file (MRCONSO.RRF). There is exactly one row in this file for each unique string or concept name within each source vocabulary. Every string or concept name in the Metathesaurus appears in this file, connected to its language, source vocabularies, and its concept identifier. Therefore we used MRCONSO.RRF to translate the variable length window (semi phrase).

MeSH descriptors in different translations get the same Concept Unique Identifier (CUI) in UMLS. For instance, the term 'breast cancer' has a Source Descriptor Unique Identifier (SDUI) 'D009203' in both English and French MeSH. D009203 is assigned to the CUI 'C0006142'. A search for the code D009203 in the UMLSKS will pull up all of the equivalent MeSH translation terms, which are all assigned to the same CUI C0027051. In other words, translation using UMLSKS depends on meaning matching. The translation process worked as follows:

- Create two dictionaries. The first one consists of all Concept Unique Identifiers (CUI) and their English equivalent phrase(s), and the second dictionary contains all MeSH descriptor identifiers and their equivalent English MeSH phrases.
- Search In MRCONSO.RRF for the French window in phrase field; since semi phrases in MRCONSO.RRF may exist in one or more records, we get all related CUI(s) for this window and their equivalent French sub phrases.
- Calculate Simple Matching Coefficient (SMC) for each French sub phrase. SMC is calculated using the following formula:

  SMC= number of matching attributes values / number of attributes
- Where attributes are the terms, if SMC equals one that means the translation has the same length as the phrase. Calculating SMC gives the most nearest meaning since it will take the shortest translation and will decrease number of unrelated terms.
- Use the CUI for the highest SMC value to get the CUI equivalent English translation. In few cases, concepts were found without equivalent English phrases for this concept. In such a case source unique identifier used to locate MeSh equivalent terms in the target language (English).
- If translation process failed, reconstruct new movable windows and start over.
- Finally weh query end reached, and successful translation achieved, translated segmented are gathered to construct new bag of words query. But if the original query contains French abbreviation wither it is translated or not it removed and replaced with its equivalent English abbreviation, e.g. IRM will be MRI instead of Magnetic Resonance Imaging.

The French stopword list used in translation experiments is downloaded from IR Multilingual Resources at UniNE [15]. This stopword list consists of 463 words, we added common terms to this stopword list to such as 'image','montrant' (means show me in English) and 'photo'.

3 Ad-hoc Experiments and Results

For ad-hoc classical retrieval task, we submitted 6 runs. For all runs we used Lemur toolkit for indexing and retrieval; Okapi retrieval model and updated stopword list. For multilingual runs, we used updated French stopword list. Best obtained results when image caption and image title only used in retrieval.

Table 1. ISSR Ad-hoc retrieval results,<P>:added paragraphs,<S> :added sentence, MAP: Mean Average Precision, R-Prec: R-Precision.

#	Run Name	Language	<P>	<S>	MAP	R-Prec	Recall
1	issr_CT	English	No	No	**0.2583**	0.2547	0.761762
2	issr_CTS	English	No	Yes	0.231	0.2388	0.761762
3	issr_CTP	English	Yes	No	0.2199	0.2564	0.760761

As shown in table1, Adding only one sentence to image textual file gives better results than adding the entire paragraph(s) from MAP point of view. It is obvious that adding extra text to image textual file increase noisy terms, thus MAP decreased.

4 Cross language retrieval Experiments and Results

Concerning multilingual retrieval ask, best results are achieved when Google online Machine Translation was used. This is not a surprise for us since we test the same techniques on 2009 dataset, but the very low MAP for all multilingual runs was not expected.

Table 2. ISSR multilingual experiments.

#	Run Name	Language	MAP	R-Prec	Recall
1	issr_CT_T_MT	French	0.1472	0.1618	0.608156
2	issr_CT_T_dic	French	0.1394	0.1145	0.453901
3	issr_CT_T_MeSh	French	0.0985	0.1186	0.596597

The main weakness in using medical dictionary is the unavailability of machine-readable dictionary (queries are translated manually).Inherit dictionary based problems, terms might not be found (dictionary coverage). For instance for query 11 '*Tous types d'images avec artériosclérose*'; none of its terms found in the dictionary.

Translation via UMLS main advantage is: the window (sub phrase) is likely to be identified and translated. However, using UMLS adds too many terms to the query. Terms that have the same concept identifier number in UMLS Metathesaurus. For that, using the UMLS adds many extra terms to English query, even if UMLS resources is eliminated to English and French MeSh only. In addition, moving the window until translation succeeds may cause loosing the translation of some terms if they were in a previously constructed window. For that, the algorithm should be modified to handle such cases: if query sentence ended and we have successful translations and failed ones too. Non translated windows may be translated by applying the moving window technique for this non translated window only.

5 Case-based Esxpermints and Results

Regarding case-based task we submitted 3 runs. As shown in table (4), almost all runs have the same performance. For that we run two paired t-test, thus to examine if the difference is statistically significant.

Table 3. ISSR case-based results.

#	Run Name	MAP	R-Prec
1	ISSR_cb_cts	0.1986	0.2292
2	ISSR_cp_ctp	0.1977	0.2253
3	ISSR_CB_CT	0.1977	0.2253

Table 4. P-value results for collections CT and CTP

Query	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25
p-value	0.3	0.9	0.4	0.3	0	0.1	0	0	1

Table 5. P-value results for collections CT and CTS

Query	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25
p-value	0.8	0.9	0.7	0.8	0.001	0.18	0	1	1

Our Null hypothesis is that: using image caption, article title and added sentence or added paragraph will give greater similarity than using image caption and article title only. Results of two paired t-tests show that this hypothesis is rejected and no significant difference between the three collections.

6 Conclusion

Submitted monolingual experiments show that ad-hoc retrieval task yielded moderate results. For case-based retrieval, adding extra images relevant terms to captions and article title did not make significant change in retrieval performance. For multilingual task, submitted experiments performance between the used three methods is close, our proposed semantic translation method via UMLS needs more investigation to eliminate adding many terms as well as continue translating while there still non translated sub phrases or windows.

Acknowledgments. We acknowledge Kareem Darwish from Cairo Microsoft Innovation Center for suggesting applying paired two t-test.

References

[1] Müller H, Clough P, et al, Evaluation Axes for Medical Image Retrieval Systems - The ImageCLEF Experience, ACM Int. Conf. on Multimedia, Singapore, November 2005.
[2] imageclef. imageclef. [Online] March 5, 2010. [Cited: July 1, 2008.] http://www.imageclef.org/2008/medical.
[3]Hersh,W., KalpathyJ., Jensen,C.: Medical Image Retrieval and Automated Annotation:OHSU at ImageCLEF 2006. In Working Notes of the 2006 CLEF Workshop,2007.
[4] Gobeill, J., Douglas, T., Patsche, E., Ruch P.: Taking Benefit of Query and Document Expansion using MeSH descriptors in Medical ImageCLEF 2009. In Working Notes of the 2009 CLEF Workshop,2009.
[5] Hersh,W., KalpathyJ., Jensen,C.: Medical Image Retrieval and Automated Annotation:OHSU at ImageCLEF 2006. In Working Notes of the 2006 CLEF Workshop, 2007.
[6] Daumke,D., Paetzold,J., Markó,P.: The Effect of Subwords on Biomedical IR. Radiograph Annotation Using Local Relational, Morphosaurus in ImageCLEF 2006.
[7] Jeffery R. Jensen, William R. Hersh:,Manual Query Modification and Automated Translation to Improve. Cross-Language Medical Image Retrieval. CLEF 2005.
[8] LEMUR toolkit website, http://sourceforge.net/projects/Lemur/
[9] Arafa, W. Ibrahim, R.,: Medical Image Retrieval: ISSR at CLEF 2009. In Working Notes of the 2009 CLEF Workshop,2009.
[10] Cognitive Computation Group at the Department of Computer Science, University of Illinois at Urbana-Champaign. http://l2r.cs.uiuc.edu/~cogcomp/atool.php?tkey=SS (accessed March 2010).
[11] Unified Medical Language System (UMLS). http://www.nlm.nih.gov (accessed March 2010).
[12] Och, F. J. Statistical machine translation: Foundations and recent advances. Tutorial at MT Summit X. (2005).
[13] Reverso Specialized dictionaries. http://dictionary.reverso.net/medical-french-english/ (last accessed at June 2010).
[14] ImageCLEF 2010. http://www.imageclef.org/2010/medical (accessed Augest 2010).
[15] IR Multilingual Resources at UniNE,http://members.unine.ch/jacques.savoy/clef/index.html, April 2010.
[16] P. Clough. Caption and query translation for cross-language image retrieval. In CLEF, pages 614{625. 2005.
[17] Pang-Ning Tan, P.N.; Steinbach, M.; Kumar, V.: “Introduction to Data Mining”., Addison Wesley (2005).
[18] Besançon, R., Millet, C.:Merging Results from Different Media: Lic2m Experiments at. ImageCLEF 2005, Cross Language Evaluation Forum 2005 Workshop, Vienna.