

Text-only Cross-language image search at medical ImageCLEF 2008

Julien Gobeill, Patrick Ruch, Xin Zhou
University and Hospitals of Geneva, Switzerland
julien.gobeill@sim.hcuge.ch

Abstract

We report on simple textual strategies with thesaural resources in order to perform document and query translation for cross-language information retrieval in a collection of annotated medical images. The keystone of our strategy for the previous medical ImageCLEF was to enrich documents and queries with Medical Subject Headings (MeSH) terms extracted from them, in order to translate the more important concepts into an intermediate language. The core technical component of our cross-language search engine is an automatic text categorizer, which associates a set of MeSH terms to any input text, with a top precision at above 90%. Nevertheless, in the new 2008 collection, images are given with more verbose captions, and with an associated article relative to a specific case study. Therefore, our strategy to enrich each document is either to collect MeSH terms from the associated article, either to extract them from the caption. Our results are fair, as we stand on the first part of the participants (0.176 for mean average precision). Nevertheless, it appears that MeSH terms collected from the relative article are not always relevant, as this article can concern a huge set of images in general, and can not to describe precisely the associated image. Moreover, the MeSH terms directly extracted from the captions lead to worst performances, possibly due to the more verbose captions. We try different strategies on weighting scheme or retrieval on articles, but without significant improvements. In conclusion, a mixed strategy to combine the two origins of the MeSH terms should be planned for the next ImageCLEF, while better performances should be obtained in the future by tuning the system with the existing benchmark.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software; H.3.7 Digital Libraries; H.2.3 [Database Management]: Languages - Query Languages

General terms

Measurement, Performance, Experimentation

Keywords

Image Retrieval, Text categorization, multimodal retrieval

1 Introduction

Cross-Language Information Retrieval (CLIR) is increasingly relevant as network-based resources become commonplace. In the medical domain it is of strategic importance in order to fill the gap between clinical records, written in national languages and research reports massively written in English. Images are also getting increasingly important and varied in the medical domain, and they become available in digital form. Despite the fact that images are language-independent, they are most often accompanied by textual notes in various languages and these textual notes can strongly improve retrieval quality (1).

Historically, the most traditional approach to IR in general and to multilingual retrieval in particular, uses a controlled vocabulary for indexing and retrieval. In this approach, a librarian selects for each document a few descriptors taken from a closed list of authorized terms. A good example of such a human indexing is found in the MEDLINE database, where records are manually annotated with Medical Subject Headings (MeSH). The MeSH is a terminology maintained by the National Library of Medicine and which exists in a dozen languages. However, it can be difficult for users to think in terms of a controlled vocabulary. Actually, the use of terminology-based systems – like most Boolean-supported engines – is often performed by professionals rather

than general users. Therefore, it can be more efficient for realistic search engine to automatically handle the documents enrichment and query expansion by MeSH concepts.

The Cross Language Evaluation Forum (CLEF) is a challenge which occurs each year since 2000. The goal of this challenge is to evaluate the participants on a common multilingual task, to establish a state of the art of the techniques used in a domain, and to build a benchmark for future evaluations. Medical ImageCLEF has started in 2004 with the goal to retrieve relevant medical images in a multilingual document collection, using visual features – images – or textual features – associated captions, titles and articles.

Our group is specialized in Natural Language Processing; nevertheless, we always participate to medical ImageCLEF, applying textual strategies based on the picture's metadata, and the use of MeSH as an intermediate language (2) (3).

2 Data and Strategies

In 2008, the collection is entirely new, as organizers were able to obtain images from the GoldMiner system (4). The collections used in the previous three medical ImageCLEF – 2005 to 2007 – were merged into one single new collection, in order to build a unique benchmark. Therefore, the 2008 ImageCLEF collection consists of new images from two radiology journals, along with their captions, article titles, and linkage to PubMed and the full text of the associated article. It contains a set of 67 115 images. In addition to the images, XML files are distributed, which contains the metadata. A detailed description of the protocol can be found in (5).

```
<Record>
  <figureID>27979</figureID>
  <figureURL>http://radiology.rsnaajnl.org/cgi/content/full/210/1/11/F1</figureURL>
  <caption>&quot;Illustration of a neonate at autopsy whose demise was attributed to thymic
death. The caption drew attention to the enormous size of the thymus, which is actually normal
in appearance. (Reprinted, with permission, from reference 6.)&quot;</caption>
  <title>The right place at the wrong time: historical perspective of the relation of the
thymus gland and pediatric radiology</title>
  <pmid>9885579</pmid>
  <articleURL>http://radiology.rsnaajnl.org/cgi/content/full/210/1/11</articleURL>
<imageURL>http://radiology.rsnaajnl.org/content/vol210/issue1/images/large/r99ja45glx.jpeg</im
ageURL>
  <imageLocalName>r99ja45glx.jpeg</imageLocalName>
</Record>
```

Figure 1: example of the metadata for one of the 67115 images. The title and caption parts are directly used by our system to retrieve relevant documents, while the PubMed id (PMID) and the article's URL can be used to extract some descriptors as MeSH terms.

Several differences between this new collection and the previous ones must be noted. Firstly, the 2008 collection contains only English texts, contrary to the previous ones which also contained French and German texts. Queries are, as previous years, asked in English, French and German. Secondly, as metadata give a PubMed id (PMID) to each document, human-generated MeSH terms can be automatically collected and associated by following the link to PubMed. Thirdly, an article is provided for each document, even if a set of images belongs to the same article – there are 4961 articles for 67115 images.

The strong point of our strategy, for we participate to ImageCLEF, is focused on associating MeSH terms to any textual components – documents or queries – in order to enrich the text with language-independent descriptors, and to perform a standard Information Retrieval process. The core technical component of our cross-language search engine is an automatic text categorizer, which associates a set of MeSH terms to any input text; the precision at high ranks of this engine for MeSH terms is above 90% (6).

```
MeSH D010437 : Peptic Ulcer
Synonyms : Ulcère gastroduodénal, Gastroduodenal Ulcer, Marginal Ulcer, Ulcus
pepticum, Ulcus marginale, Ulcus gastroduodenale
```

Figure 2: example of a MeSH concept as indexed by our categorizer: behind the MeSH identifier stands the concept described with French, English and German synonyms.

We merge three versions of MeSH (English, German and French (see figure 2)) in order to enrich each document with several MeSH terms – between 3 and 8 in 2006, 15 in 2007 – and their unique identifier, making them efficient regardless of the original language of the document (see figure 3). The number of terms is an important parameter, as in 2006, the more MeSH terms were added, the best the run was. We finally showed in 2007 that with the previous collection, the ideal number of MeSH concepts per document was around 15 (2) (3) – which is the mean for an article in MEDLINE. The enriched documents are then indexed in a standard way.

- a) 8000000310
 Upper Gastrointestinal Ulcers
 Images from the National Endoscopic Database
 Endoscopy, Gastrointestinal; Peptic Ulcer
 Location: Duodenal Bulb, Not bleeding, clear ulcer base.
- b) 0127714|peptic ulcer|D010437
 0043948|endoscopy|D004724
 0039258|endoscopy, gastrointestinal|D016099
 0043827|bleeding|D006470
- c) Upper Gastrointestinal Ulcers
 Images from the National Endoscopic Database
 Endoscopy, Gastrointestinal; Peptic Ulcer
 Location: Duodenal Bulb, Not bleeding, clear ulcer base.
 peptic ulcer D010437
 endoscopy D004724
 endoscopy, gastrointestinal D016099
 bleeding D006470

Figure 3: example of the MeSH categorization for one document in previous ImageCLEF. a) Caption of an image of the endoscopic collection (in English). b) output of our MeSH categorizer for the previous caption : concept score in first column, concept in second one, MeSH identifier for the third one. c) final document indexed : the caption is enriched with found MeSH descriptors and their MeSH id, making them language-independent.

The same MeSH categorization is then performed on queries; according to past studies, the ideal number of terms associated by each query is 3 (2). For instance, if a German query deals with magen-darm-endoskopie (see figure 4), this concept has great chances to be mapped by our categorizer, and to enrich the query with this German form and the MeSH id: then, even if we work in an English collection, the MeSH id D016099 will be a strongly discriminant feature. The search engine then performs a standard Information Retrieval process. So, MeSH is seen like an intermediate language between documents and queries (2).

```

« Magen-Darm-Endoskopie mit Geschwür »
Magen-Darm-Endoskopie D016099
Geschwür D010437
Endoskopie D004724
  
```

Figure 4: example of the same treatment for a German query: even if we retrieve German terms in an English collection, the MeSH identifiers will be strongly discriminant features.

With the new collection used for ImageCLEF 2008, and the PMID associated to each document, our strategy is lightly different. Human-generated MeSH terms can be collected for each document, thanks to the PMID contained in the metadata. So, even if our MeSH categorizer obtains good results, we can suppose that “official” descriptors are more accurate and more complete. So, one strong skill of our strategy becomes needless, but we choose to enrich images with the MeSH terms attached to their PMID. Nevertheless, we choose to keep a run where metadata are enriched with MeSH terms found by our categorizer. We try different weighting schemes too, and different combinations between captions, full texts and MeSH terms. The last strategy is to supply a textual run to another team of University and Hospitals of Geneva, Xin Zhou and Henning Muller, who work on visual Information Retrieval, in order to produce a mixed run. More details for each run are given in the Results part.

An important point for us is the use of the three languages into the same query. We think that having the same queries, perfectly translated into three languages, is not a realistic task. No human user asks a question into three different perfect translations in a system. So, we think preferable, for each query, to make a run for each language: English (EN), French (FR) and German (GE). This is an opinion that we defend in each ImageCLEF we participate, even if we can lose some performance, as seen in 2007 (3).

2.1 MeSH-driven Text Categorization

Automatic text categorization has been studied largely and has led to an impressive amount of papers. A partial list of machine learning approaches applied to text categorization includes naïve Bayes (7), k-nearest neighbours (8), boosting (9), and rule-learning algorithms (10). However, most of these studies apply text classification to a small set of classes; usually a few hundred, as in the Reuters collection (11). In comparison to this our system is designed to handle large class sets (12): retrieval tools used are only limited by the size of the inverted file, but 10^{5-6} documents is still a modest range. Our approach is data-poor because it only demands a small collection of annotated texts for fine tuning: instead of inducing a complex model using large training data, our categorizer indexes the collection of MeSH terms as if they were documents and then it treats the input as if it was a query to be ranked regarding each MeSH term. The classifier is tuned by using English abstracts and English MeSH terms. Then, we apply the system on the medical ImageCLEF collection. For tuning the categorizer, the top 15 returned terms are selected because it is the average number of MeSH terms per abstract in the OHSUMED collection. When applied on the medical ImageCLEF collection, the number of categories to be attached to every document will be an important parameter.

2.2 Collection and Metrics

The mean average precision (map): is the main measure for evaluating ad hoc retrieval tasks (for both monolingual and bilingual runs). Following (13), we also use this measure to tune the automatic text categorization system. We tune the categorization system on a small set of OHSUMED abstracts: 1200 randomly selected abstracts were used to select the weighting parameters of the vector space classifier and the best combination of these parameters with the regular expression-based classifier.

3 Methods

Two main modules constitute the skeleton of our categorization system: the regular expression (RegEx) component, and the vector space (VS) component. Each of the basic classifiers implements known approaches to document retrieval. The first tool is based on a regular expression pattern matcher (14), it is expected to perform well when applied on very short documents such as keywords: MeSH terms do not contain more than 5 tokens. The second classifier is based on a vector space engine. This second tool is expected to provide high recall in contrast to the regular expression-based tool, which should privilege precision. The former component uses tokens as indexing units and can be merged with a thesaurus, while the latter uses stems (Porter).

Regular expressions and MeSH thesaurus. The regular expression search tool is applied on the canonic MeSH collection augmented with the MeSH thesaurus (120'020 synonyms). In this system, string normalization is mainly performed by the MeSH terminological resources when the thesaurus is used. Indeed, the MeSH provides a large set of related terms, which are mapped to a unique MeSH representative in the canonic collection. The related terms gather morpho-syntactic variants, strict synonyms, and a last class of related terms, which mixes up generic and specific terms. The system cuts the abstract into 5-token-long phrases and moves the window through the abstract: the edit-distance is computed between each of these 5 token sequences and each MeSH term. Basically, the manually crafted finite-state automata allow two insertions or one deletion within a MeSH term, and ranks the proposed candidate terms based on these basic edit operations: insertion costs 1, while deletion costs 2. The resulting pattern matcher behaves like a term proximity scoring system (15), but is restricted to a 5-token matching window.

Vector space classifier. The vector space module is based on a general IR engine with the tf.idf weighting schema. The engine uses a list of 544 stop words. As for setting the weighting factors, we observed that cosine normalization was especially effective for our task. This is not surprising, considering the fact that cosine normalization performs well when documents have a similar length (16).

Classifier fusion. The hybrid system combines the regular expression classifier with the vector-space classifier. We do not merge our classifiers by linear combination, because the RegEx module does not return a scoring consistent with the vector space system. The combination does not use the RegEx's edit distance, and instead it uses the list returned by the vector space module as a reference list, while the list returned by the regular expression module is used as boosting list, which serves to improve the ranking of terms listed in RL.

3.1 Cross-Language Categorization and Indexing

To translate the medical ImageCLEF textual contents (queries or documents), we transform the English MeSH mapping tool described above, attributing MeSH terms to English abstracts or queries. Thus, the English, French, and German version of the MeSH are simply merged in the categorizer. We use the weighting schema and system combination ([dtu.dtn | ltc.atn] + RegEx) as described in (2). Then, the annotated collection is indexed using the vector-space engine used by the categorizer. For the document indexing, we rely on weighting schemas based on pivoted normalization (dtu.dtn): because the documents have a very variable length in the collection such a factor can be important. A slightly modified version, ltc.atn, which has shown some effectiveness for the TREC Genomics, is used too in a run. The English stop word list is merged with a French and a German stop words list. Porter stemming is used for all documents.

4 Results and Discussion

We then describe each run separately.

4.1 Baseline run

The Baseline run (BL) is submitted to evaluate the performance of our Information Retrieval Engine alone. The strategy is simple: for each image, the caption and the title are indexed. Then, queries are submitted in 3 languages, without add of MeSH descriptors.

map	BL
EN	0.136
FR	0.069
GE	0.07

Table 1: mean average precision (map) for the Baseline run (BL)

As captions and titles are in English, it is not stunning that the English run is the best one. Nevertheless, map of French and German runs, without any strategy, is quite high. Actually, map is very different depending on the queries, because some German and French queries have very specific disease or anatomic names, which are unchanged across languages. For example, the query 24 deals with “malformation de Budd-Chiari” in French. “Budd-Chiari” is a very specific term which is supposed to be a strongly discriminant feature: the query 24 has a map of 0.89 for French. On the contrary, the query 21 “photographies de tumeur” in French has no chance to be associated with a relevant English word contained in a caption: map for the query 21 is 0.002.

4.2 MeSH run

The MeSH run (MH) is finally the best one, while it uses the simplest strategy. For each image, the caption, the title and the MeSH terms extracted from MEDLINE with their PMID – MeSH term + MeSH id as seen in figure 3 – are indexed. Queries are enriched by 3 MeSH terms too (see figure 4), and are then submitted.

map	MH
EN	0.176
FR	0.105
GE	0.076

Table 2: mean average precision (map) for the MeSH run (MH)

The benefit for English is +30%, for French +52%, and for German +8%. This is nearly the same order of performance than in the previous ImageCLEF for our strategy. But while we thought that MeSH terms collected from PubMed will be more accurate and more complete, it's not always the case. For example, the relevant document g01oc14g23x deals with "Budd-Chiari syndrome". But the MeSH concept "Budd-Chiari syndrome" (D006502) is not a MeSH term for the corresponding PMID (11598252). While the query 24 is enriched with the concept "Budd-Chiari syndrome D006502", the relevant document is not; so we lose the discriminant power of the feature D006502, which is the keystone of our strategy.

The associated article – and its MeSH terms – seems to be sometimes too general; it perhaps describes more a case study, through a set of images, than the specific image. For example, the PMID 11598252 corresponds to 46 images and deals with hepatocellular carcinoma; the image showing a Budd-Chiari syndrome is not necessarily relevant with this general set, and with the associated MeSH terms.

4.3 Assignment of MeSH terms for German queries with OVID

A problem that we manually discover with the German queries is the complexity of the words used, especially when several words are aggregated in only one. For instance, for the German query 25 "Merkelzellkarzinom", our categorizer returns no MeSH terms, while we can suppose that some concepts can be mapped if the word was split in "Merkel" "Zell" and "Karzinom". As the query is enriched with no MeSH terms, the retrieval step returns no documents.

A simple strategy that we choose to beat this difficulty is to use OVID. OVID is a multilingual search engine in MEDLINE, developed and maintained by the University and Hospitals of Geneva (17). Queries which return no MeSH terms with our MeSH categorizer are submitted to OVID. We obtain a list of relevant documents from PubMed. Then, we simply retain the 3 most frequent MeSH terms in the 10 most relevant documents, and we enrich the query with these MeSH terms.

map	MHnOVID
GE	0.107

Table 3: mean average precision (map) for the OVID strategy (MHnOVID)

The benefit for German is +40%. We think it's a better strategy, at least for our approach based on MeSH descriptors, than trying to translate the German query with an automatic translator – as Babelfish – because we translate directly the query into the chosen intermediate language, i.e. the MeSH. Nevertheless, it could be interesting to compare these results with runs composed with translations strategies.

All following runs are performed with German queries enriched with this strategy.

4.4 ltc run

The ltc run is the same as the MeSH run, but the weighting scheme used for Information Retrieval is ltc.atn instead of dtu.dtn. ltc.atn is a weighting scheme which showed good performances at TREC Genomics last years. See 3.1 for more details.

map	ltc
EN	0.155
FR	0.106
GE	0.098

Table 4: mean average precision (map) for the ltc run (ltc)

The utilization of this weighting scheme shows no improvements for English and French, while it lightly increase German map.

4.5 weight mix run

The weight mix run (mixWeight) is computed by linearly combining the two weighting schemes: dtu.dtn and ltc.atn. Document's scores from two runs are normalized, then merged.

map	mixWeight
EN	0.17
FR	0.115
GE	0.103

Table 5: mean average precision (map) for the weight mix run (mixWeight)

While the linear combination increases the map for German and French, it doesn't for English, which brings the best results. It seems that the combination should be obtained with different factors than 50-50; nevertheless, these factors have to be tuned, and it was impossible to do before having a benchmark.

4.6 mix papers run

The mix papers run (mixPapers) is the more sophisticated one. We start working from the MH run. All scores are normalized. Then, we index the full texts attached to the images, and perform a retrieval from the query on this collection: the top relevant documents are then used in order to boost the associated images in the first run (score increased by 10%).

map	mixPapers
EN	0.15
FR	0.077
GE	0.118

Table 6: mean average precision (map) for the mix papers run (mixPapers)

There is no improvement obtained for English, but we obtain our best run for German. We can suppose that, as the MeSH strategy is less effective with German (see 4.2) and captions are relatively short documents, articles bring more textual data in order to compare with the German queries. Nevertheless, for English, articles seem to bring more noise than more precisions; there still needs to work on the coefficients of the combination in order to confirm this hypothesis.

4.7 MeSH terms extracted from captions

For this run, the MeSH terms associated to captions and title are not extracted from MEDLINE thanks to their PMID, but they are mapped with our MeSH categorizer. We choose to keep 15 MeSH terms per document.

map	capMH
EN	0.134
FR	0.093
GE	0.073

Table 7: mean average precision (map) for the captions MeSH run (capMH)

The improvement for English is null. Nevertheless, the improvement is significant for French (+34%). It appears that while the MeSH terms collected from the associated article are not relevant for each image, the MeSH terms extracted with our categorizer are certainly not precise enough. We can suppose that the more verbose captions of the images, compared to the previous years' ones, leads to poorer performances of our MeSH categorizer. The solution could be to mix the two origins of MeSH terms in order to obtain a more complete set of descriptors for each image.

4.8 map of English MeSH run depending on queries type

We split the results of our best run, MeSH run for English (see 4.2) in the three types of queries: visual (1-10), mixed (11-20) and semantics (21-30).

map	MH-EN
visual	0.091
mixed	0.094
semantics	0.344

Table 8: mean average precision (map) for the English MeSH run (MH-EN) depending of queries type

This is obviously the semantics queries which achieved the best results. Nevertheless, we notice that this run is nearly at the same rank compared to all participants' runs, no matter we split the queries in visual, mixed or semantics ones. Actually, it appears that the teams which choose a visual strategy obtain poor results on this benchmark. Taking a closer look to participants' results shows that some textual runs performs better in the visual queries than in the mix ones (SINAI-sinai_CT_Mesh for instance), and some visual runs performs better in the mixed queries than in the visual ones (GE-GE_GIFT8). We suppose that these anomalies should disappear when the visual techniques will be more adapted to this collection.

4.9 combination of the best run with a visual run from another participating team

To obtain these last runs, we supply our best run (MH-EN) to another team of University and Hospitals of Geneva, Xin Zhou and Henning Muller, who are specialized in Visual Information Retrieval.

Map	GE-GE_GIFT8	GE-GE_GIFT8_EN0.5
Visual	0.015	0.076
Mixed	0.061	0.117
Semantics	0.027	0.061
all	0.035	0.085

Table 9: mean average precision (map) for the best run of the Xin Zhou and Henning Muller team, and for the best mixed run (GE-GE_GIFT8_EN0.5)

X Zhou and H Muller performed runs relied mainly on GIFT (3); their runs are more supposed to be baseline runs than candidates for the high ranks. As, moreover, the visual strategies leads quite poor

performances this year, it is not surprising that the combination with a visual run leads to no improvements. We hope this combination will be better for medical ImageCLEF 2009.

5 Conclusion and Future Work

The keystone of our strategy for the previous medical ImageCLEF was to enrich documents and queries with Medical Subject Headings (MeSH) terms extracted from them, in order to translate the more important concepts into an intermediate language. Nevertheless, in the new 2008 collection, images are given with more verbose captions, and with an associated article relative to a specific case study. It appears that MeSH terms of the associated article collected from MEDLINE are not always relevant, as this article can concern a huge set of images dealing with a more general subject, and can not to describe precisely the associated image. Moreover, the MeSH terms extracted directly from the captions leads worst performance, possibly due to the more verbose captions.

For the future ImageCLEF, a mixed strategy to combine the two origins of the MeSH terms should be planned. Better performances should be obtained too by tuning the system with the existing benchmark.

References

1. *A review of content-based image retrieval systems in medicine - clinical benefits and future directions.* **H Müller, N Michoux, D Bandon and A Geissbuhler.** 2004, International Journal of Medical Informatics, pp. 73:1-23.
2. *Query and Document Translation by Automatic Text Categorization: A Simple Approach to Establish a String Textual Baseline for ImageCLEFmed 2006.* **J Gobeill, H Muller and P Ruch.** 2006. ImageCLEF.
3. *University and Hospitals of Geneva at ImageCLEF 2007.* **X Zhou, J Gobeill, P Ruch and H Muller.** CLEF 2007 Working notes.
4. *GoldMiner: a radiology image search engine.* **C, Kahn CE Jr and Thao.** 188, 2007, American Journal of Roentgenology, pp. 1475-1478.
5. [Online] <http://ir.ohsu.edu/image/2008protocol.html>.
6. *Automatic Assignment of Biomedical Categories: Toward a Generic Approach.* **Ruch, P.** 22(6), 2006, Bioinformatics, pp. 658-64.
7. *A comparison of event models for naive bayes text classification.* **Nigam, A McCallum and K.** 1998.
8. *An evaluation of statistical approaches to text categorization.* **Yan, Y.** 1999, Journal of Information Retrieval, Vol. 1, pp. 67-88.
9. *BossTexter: A boosting-based system for text categorization.* **Singer, R Shapire and Y.** 2000, Machine Learning, pp. 135-168.
10. *Automated learning of decision rules for text categorization.* **C Apte, F Damerau and S Weiss.** 1994, ACM Transactions on Information Systems (TOIS), Vol. 12, pp. 233-251.
11. *A system for content-based indexing of a database of news stories.* **Weinstein, P Hayes and S.** 1990, Proceedings of the Second Annual Conference on Innovative Applications of Intelligence.
12. *Learning-Free Text Categorization.* **P Ruch, R Baud, and A Geissbuhler.** 2003, LNAI 2780, pp. 199-208.
13. *Combining classifiers in text categorization.* **Croft, L Larkey and W.** 1996, SIGIR, ACM Press, New York, pp. 289-297.
14. *A tool to search through entire file systems.* **Wu, U Mamber and S.** Proceedings of the USENIX Winter 1994 Technical Conference, San Francisco, pp. 23-32.
15. *Term proximity scoring for keyword-based retrieval systems.* **Savoy, Y Rasolof and J.** ECIR 2003, pp. 101-116.
16. *Cross-language information retrieval (CLIR) track overview.* **Sheridan, P Schäuble and P.** 1998, Proceedings of The Sixth Text Retrieval Conference (TREC6).
17. [Online] <http://gateway.ovid.com/>.