Brazilian Journal of Information Science

A MULTILINGUAL AND LOCATION EVALUATION OF SEARCH ENGINES FOR WEBSITES AND SEARCHED FOR KEYWORDS

Anas AlSobh
Ahmed Al Oroud
Mohammed N. Al-Kabi
Izzat AlSmadi

Yarmouk University
Jordan

ABSTRACT

Search engines are competing to be seen as universal, consistent and language independent. In principle, users searching for information through the Internet should get consistent information regardless of the language of the words they are searching for and regardless of the language of the matching or the relevant documents. Nevertheless, the language should affect the sequence or the ranking of the retrieved results. In this project, several tools are built to evaluate words and statements in several languages. Results are evaluated and compared for possible correlation. Another tool is built to crawl websites from different languages and locations in order to measure several aspects of those websites. Results from both studies showed that while it seems that popular search engines are making very good progress toward building search engines that are language and location independent, however, there are some limitations and situations where search for results can be biased toward the popularity of the website language and/or location.

Keywords: Information Retrieval; Search Engines; Natural Language Processing; Translation; Text Matching; Language Queries.

INTRODUCTION

Information overloading is a continuous concern for information retrieval researchers. Users in many cases are overwhelmed by the amount if information retrieved as a result of their query on a search engine. Internet users all over the world are surfing the Internet looking for information relative to their interests. Their main tool to look for such information is the search engines. Search engines keep their database updated by continuously crawling through the Internet to collect and index all WebPages, documents and contents of Websites. In order to increase their

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

popularity, large search engines are continuously evolving to cover services in international languages. For example, Google launches recently several new local domains and services for several languages. For example, in the case of the Arabic language, there are new services that include: Google Translate, Suggest, and Google Ejabat for answering Arabic questions, auto completion, Google zeitgeist, Translated Search, Tashkeel, etc.

Google as the primary search engine for many users around the world is continuously evolving, improving and expanding the website tools to cover different utilities and to target users all around the world using different languages. Users of languages other than English can search using the English keywords or they can search using their own language. Retrieved results may not exactly match. This can be justified through saying that the user who search using a keyword in a particular language is interested to get first results in that specific language. Users also who search from a specific location may want to get first results from their own country or area as more relevant pages than those of other languages or other continents. However, in both cases, eventually results should come to the same, or nearly same, results. Search engine indexers should isolate the layer of the location and the language from the actual content and documents retrieved and indexed in their own library.

The ultimate goal of this research is to propose building indexers that are language independent. We will evaluate Google translate along with several other open source dictionaries such as WordNet (wordnet.princeton.edu) etc to compare search retrieved results between Arabic words and their related English ones.

2 RELATED WORK

Salton (1969) refers to Cross-Language Information Retrieval (CLIR) in the late 60s of the 20th century, where a multi-lingual thesaurus is used for the documents and the queries. Salton asserts that CLIR could be effective as the mono-lingual information retrieval.

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

Al-Onaizan et al. (2002) study presents a solution to the problem of translating named entity phrases. This is a difficult problem, since these phrases are in most cases specialty related phrases not general ones. As a result, you could not find them in general dictionaries. They tackled the problem of translating named entity phrases by presenting a novel algorithm dedicated to translate Arabic to English. This novel algorithm adopts different approaches for different types of named entity phrases, where the translation process is based on two main steps. In the first step a ranked list of candidate translations are produced. In the second step the candidate translations are rescored depending on monolingual clues. Later on, candidate translations in the first list are re-ranked. Their algorithms transliterate and translate Arabic words to English, and then determine whether to use transliterated or translated English terms.

There are attempts to enhance the queries entered into search engine boxes, and one of these attempts by Loia et al. (2007). It was based on adding semantically similar queries to the original query. Submitting the original query in addition to the semantically equivalent queries to the search engine yields different results. The results are then unified into one filtered list. This approach aims to help the users to formulate their queries during search session, besides getting better results.

Different query similarity approaches have been evaluated by (BALFE et al., 2005). Those approaches are classified into three categories: term based similarity the results based metrics and the behavior of users in selecting relevant pages. The hit matrix representation is used to find the common terms between different queries. The relevancy score was used as a measure in a results-based approach. The selection based approaches used the user selection measures for relevant pages to find the similarity between queries. The results indicated that term based approaches achieve the best results in terms of precision and recall.

A graph to visualize the related queries which utilizes hybrid query similarity measure to generate query clusters for each submitted query was proposed by (LIN et al., 2004). The graph is generated by applying query clustering algorithms and TF-IDF algorithm to build the query repository, upon which the query clusters are built. A questionnaire was used to measure users' satisfaction of the new approach for

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

related queries suggestion. The results indicated that in terms of time, users spent less time in formulating their queries using the graph based method. Guo and Bian (2008) proposed a multi-lingual information retrieval system for patent documents in English and Japanese. Different web translators such as Google and Excite translators are used to translate queries. The language-independent indexing technology is used to process the text collections in various Asian languages. The results indicated that the proposed method achieved effective results. However, the proposed information retrieval system was not a web-based one. In addition, no relevant feedback procedure was used.

Lianhau et al. (2009) build a multilingual information retrieval system called MARS. The creation of MARS is based on manipulating a collection of documents into clusters of comparable sets by finding underlying associations between them. Clustering of documents was performed offline; the clustering was actually a basis for retrieving a comparable, multilingual and related document online according to user queries. MARS supported only simple queries in GUI. It was less appropriated for complex ones.

The effectiveness of a multi-lingual Information Retrieval System (IRS) capable to deal with four languages; English, Chinese, Japanese and Korean was evaluated by Savoy (2005). A combined translation approach was used; where the results indicate that the combined translation strategy seemed to enhance retrieval effectiveness for Chinese and Japanese, but not for Korean. This study also addressed the merging strategy of results sets generated in different languages, where the Z-score merging procedure achieved about 5% better performance than the traditional round-robin approach.

An ontology driven cross language information retrieval was described by Nilsson et al. (2006). A domain specific query expansion and translation was used. The process of building system ontology was composed from collecting concepts specific to the university, for the purpose of query expansion. Synonyms and hyponyms were used. The corresponding terms in the target language were used for the cross language search. The system was evaluated by users. However the proposed system has some deficiencies in the translation module.

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

Jang et al. (2002) proposed a cross language information retrieval performance on a set of Korean queries for English and Chinese documents. A dictionary based translation method was used. A bilingual dictionary was used for query translation. An ambiguity resolution technique was used to remove un-necessary terms as well as unuseful words that have no effects on retrieval performance. For Korean English queries, the performance of the system was successful. However, for Korean-Chinese, the system performance was low. It was found also from the results that bilingual translation has its own problems and results in low performance.

A statistical based approach was used for query translation by Christof et al. (2005). A bilingual dictionary as well as a monolingual corpus was used in the experiments. An algorithm is proposed that combines the term association measures with iterative machine learning for probability calculation. The learned translation probabilities are used as query term weights and were also integrated into a vector space retrieval system. The results have shown that involving an incremental approach for query translation may results in a better performance for cross-language information retrieval.

Graph theory and the pattern based method are one of the proposed techniques used by the researchers to resolve queries translation ambiguities in CLIR systems. Zhou et al. (2008) proposed an enhanced hybrid graph-pattern based approach to improve the query translation performance in the cross language information retrieval. The proposed method starts translating candidate terms from a bilingual dictionary. Hence several translations may exist for the same term. A pattern matcher is used in the second step for unknown and ambiguous terms. Then all translations that were generated in the first step were forwarded to a graph representation, where terms co-occurrences were used to get the best translation sequence. The evaluation results reveal a promising improvement relative to traditional methods.

The syntax relationship between a set of related words could be easily found on several search engines (e.g. Google, Yahoo). However, search engines may not consider the semantic relationships that may exist between concepts. Danushka et

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

al. (2009) proposed a method to find the similarity between a set of semantically related terms. A lexical patterns extraction algorithm was used to represent common semantic relationships between terms (e.g. Google Acquire YouTube). A sequential patterns algorithm was also used to cluster a set of patterns in appropriate way, and then a feature vector was constructed to find the relational similarity between the extracted patterns. The test that was conducted on the proposed method reveals an enhancement in terms of performance and processing time.

In the recent years the vision of using search engines has been moved from retrieving a huge irrelevant data into retrieving useful information that can be analyzed by experts. Hence we are currently moving toward mining the pages retrieved by any Web search engine. This issue was discussed by Erinjeri et al. (2009). Google search engine was used to mine the radiology reports using free and open-source technologies. A tool called Radsearch was developed as part of their research, where it is actually built above Google infrastructure. This tool enables Google to retrieve some Web pages related to radiological reports.

Chew and Abdelali (2008) have studied the effects of language relatedness on the performance of cross language information retrieval systems. This approach is used to measure the effects of using Semitic languages within cross language information retrieval systems that include Arabic. Results of the study indicated that CLIR performance was enhanced extensively.

3 GOALS AND APPROACHES

3.1 Searched for Keywords

In order to evaluate some language related issues, we built a small database of most visited keywords in Arabic in several countries. The most visited Arabic keywords stored within the database are collected from Google, Alexa and some other websites that track information of most visited keywords per country. Those websites keep tracking Internet users' behaviors and the keywords they search for, or in other words analyzing global search volume on particular keywords. In Google, this

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

is accomplished through several tools. First, Google suggest or auto completion is a method to show users who start typing letters the most visited words that match their typed letters. The second and third sources of information for most visited words are: Google Trend (www.google.com/trends) and Google zeitgeist (http://www.google.com/intl/en-/press/zeitgeist/index.html). The fourth tool from Google is the Search-based keyword tool; Sktool (http://www.google.com/sktool) that provides keyword ideas. The fifth tool from Google is Google Insights for Search (http://www.google.com/insights/search). Using Google Translate and some other Arabic to English dictionaries, the dataset of popular words is translated into English. For both Arabic and their relevant English words, number of related words and returned search documents are returned. The total number of collected terms exceeds 6000 terms for each language. In some cases, some of those keywords are repeated. However, as those come from different countries, they kept as they are expected to show different results in terms of number of matched documents or traffic. A crawler and a robot are built to collect data automatically.

Related queries or searches related to in Google (Figure 1) show the keywords that are related to the current searched keyword(s). They are usually up to 8 results (usually show in the bottom and sometimes in the top of the page) in Google and will stay fixed on all search return pages. Such related queries may depend on several parameters. This may include the history of searches Google keeps for individuals (one, who search for x, will usually search for y also). It may also depend on natural language processing and semantics. Traffic is another factor. Keywords that appear as related searches have already toggled a filter and been promoted to that position as a result of search volume.

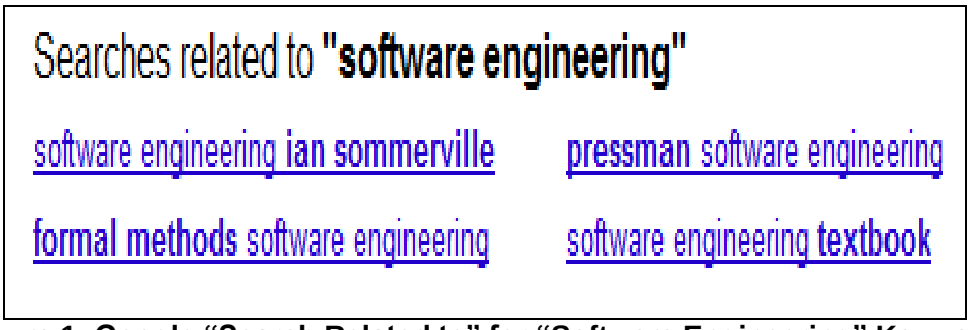

Figure 1: Google "Search Related to" for "Software Engineering" Keyword.

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

3.2 Experiments and Results

A database is built from the 6589 keywords collected in both Arabic and English. For each word, in Arabic and English, the number of returned results (i.e. the approximate returned results or documents from the search engine to refer to the number of documents that the search engine found), and the number of related queries or “search related to” keywords (i.e. the number of query terms related to the keyword(s) the user used) are collected. The goal was to study the variations and the dependency of search terms and related documents on the language used for search. Table 1 shows the related queries correlation between English and Arabic. Google returns a number between 0 and 8.

Table 1 shows that despite those are the popular keywords in Arabic, Google found more related words in their English relevant terms. However, in Table 2 number of related or retrieved documents is somewhat similar.

In order to see the other side of the picture, 1688 keywords are selected using Google Sktool, Google Trend and Alexa. Those are the most worldwide used words. Table 3 shows the number of search related words.

Table 1 - “Related queries” comparison between English and Arabic for Arabic popular keywords.

% of English keywords that return more related queries	% of Equal number of queries	% of Arabic keywords that return more related queries
41.44	36.05	22.43

Table 2 - “Retrieved documents” comparison between English and Arabic for Arabic popular keywords.

% of English keywords that return more related documents	% of Equal number of documents	% of Arabic keywords that return more related documents
51.15	0.05	48.72

Table 3 - “Related queries” comparison between English and Arabic for English popular keywords.

% of English keywords that return more related queries	% of Equal number of queries	% of Arabic keywords that return more related queries
51	39.4	8.6

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

Table 4 - "Retrieved documents" comparison between English and Arabic for English popular keywords.

% of English keywords that return more related documents	% of Equal number of documents	% of Arabic keywords that return more related documents
90.38	6.7	2.92

Table 4 shows the percentage of retrieved documents between the Arabic and English. Table 4 shows that more than 90% of the retrieved documents in English are larger than the retrieved documents in Arabic. Only less than 3% of words in Arabic retrieved more documents.

Figure 2 shows a sample of those documents where there is a large difference in each row between retrieved documents in English and their translated ones in Arabic.

English Word or Phrase	E-Retreived docs	Arabic Word or Phrase	A-Retreived docs
mortgage calculator	40,900,000	حاسبة التمويل العقاري	13,500
microsoft project	89,500,000	مشروع مايكروسوفت	765,000
ms project	285,000,000	مشروع السيدة	1,030,000
payroll	29,900,000	الرواتب	1,370,000
amortization calculator	3,080,000	آلة حاسبة اهتلاك	79
mortgage payment calculator	36,600,000	دفع الرهن العقاري حاسبة	1,430
interest calculator	45,200,000	آلة حاسبة الفائدة	23,200
content management system	867,000,000	نظام إدارة المحتوى	896,000
accounting software	35,400,000	برامج المحاسبة	1,480,000
home loan calculator	35,100,000	حاسبة قرض المنزل	49,400
project management software	94,100,000	برمجيات إدارة المشاريع	1,640,000
sales jobs	110,000,000	مبيعات وظائف	526,000
crm software	12,500,000	برمجيات إدارة علاقات العملاء	85,100
content management	202,000,000	إدارة المحتوى	4,950,000
amortization table	457,000	جدول السداد	101,000
peachtree	5,560,000	بيتر	47,700
loan payment calculator	16,300,000	تسديد القرض حاسبة	1,710
bulk email	23,500,000	البريد الإلكتروني السائبة	8,080
timesheet	1,770,000	الجدول الزمني	356,000
inventory management	28,800,000	إدارة المخزون	335,000

Figure 2: Sample "retrieved documents number" between English and Arabic keywords.

It is expected that the search engines algorithms will prioritize the retrieved documents based on several factors such as traffic or popularity. This may explain the reason that Arabic words may not retrieve documents in the same order as those

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

in English (i.e. same words translated) as this reflects the popularity of the words on a specific country or region. However, this should not affect, to a large extent, the number of the retrieved documents. Table 4 indicates that popular words in the world have very small number of retrieved documents in Arabic.

Table 5 - "Related queries" comparison between English and other languages for English popular keywords.

% of English keywords that return more related queries	% of Equal number of queries	% of German keywords that return more related queries
32.69	50.96	15.38
% of English keywords that return more related queries	% of Equal number of queries	% of French keywords that return more related queries
67.3	30.7	0
% of English keywords that return more related queries	% of Equal number of queries	% of Chinese keywords that return more related queries
48	29.8	20.2

Table 6 - "Retrieved documents" comparison between English and other languages for English popular keywords.

% of English keywords that return more related documents	% of Equal number of documents	% of German keywords that return more related documents
83.65	6.73	9.62
% of English keywords that return more related documents	% of Equal number of documents	% of French keywords that return more related documents
74	12.5	13.5
% of English keywords that return more related documents	% of Equal number of documents	% of Chinese keywords that return more related documents
89.4	0	10.6

In order to summarize, Table 7 shows search related queries and number of retrieved documents between the 5 languages. Percentages are shown relative to English (i.e. focus in this Table is only on the languages percentages relative to English).

Table 7 - Search related queries and number of retrieved documents between the different languages relative to English.

Language	Search related queries	Number of retrieved documents
Arabic	8.6	2.92
German	15.38	9.62
French	0	13.5
Chinese	20.2	10.6

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

Table 7 shows that in both “search related queries” and “number of retrieved documents” clearly indicate the English language is dominating the Internet relative to the 4 other selected languages.

There are two major roles of the language on the websites and their users. The impact of the number of (native) speakers of a certain language on the number of web-hosts in that language, and the impact of the number of web-hosts in a certain language on the number of hyperlinks linking from/between websites of that language. The number of websites and readers or viewers can both benefit from each other. The large number of existed websites through the Internet in a specific language may contribute to improving the popularity of the language. On the other hand, a language, such as English, with a large number of speakers will give better opportunity and more traffic for the websites in this language.

The linking from and to a website is also another major factor affecting the popularity of any website. This is also directly related to the language popularity and number of native speakers. In English in particular, the majority of speakers are not native and there are many websites around the world that are written in two languages: the native language and English language.

3.3 Popularity Metrics

In order to correlate the relation between the language and the country of the website from one side, with its popularity from another side. A tool is developed to calculate the inlinks and outlinks of the top 10 websites from 6 countries selected based on their language. Those 6 countries are: USA for English, Germany for German language, Spain for Spanish, China for Chinese, France for French and Egypt for Arabic. Using Alexa.com, the top 10 visited websites from those 6 countries are selected and their inlinks and outlinks were gathered. Our developed tool to measure inlinks and outlinks used several algorithms for preprocessing to decrease or illuminate many irrelevant and redundant links to websites that does not affect the collected metrics to a large extent. Example of those illuminated web pages or components are those pages that are automatically generated by web design tools

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

and hence will see it in all websites. Table 8 shows results gathered from all selected websites. Zero links in some websites indicate a rerouting of the website such as www.msn.com that is moved to www.bing.com.

Table 8 - Popularity metrics for the top 10 websites in the 6 selected countries.

USA		France	
OutLink	**Inlink**	**OutLink**	**Inlink**
56	7320	78	1247
36	3266	**159**	**2847**
159	2847	**16**	**329**
745	7485	**745**	**7485**
1728	3249	**56**	**7320**
331	2355	**159**	**2847**
369	628	421	896
770	1217	**0**	**916**
16	329	6	274
371	1120	58	737
Egypt		**China**	
OutLink	**Inlink**	**OutLink**	**Inlink**
48	27	44	5678
36	**3266**	77	765
159	**2847**	56	6112
56	**7320**	123	682
745	**7485**	**56**	**7320**
149	480	712	2516
126	687	511	1621
16	**329**	233	867
369	**628**	82	657
37	**1604**	511	1543
Spain		**Germany**	
OutLink	**Inlink**	**OutLink**	**Inlink**
32	457	392	1370
159	**2847**	**56**	**7320**
16	**329**	**159**	**2847**
56	**7320**	**745**	**7485**
745	**7485**	722	1180
770	**1217**	**1728**	**3249**
159	**2847**	1254	2311
195	1261	**159**	**2847**
151	1052	1213	3126
0	**916**	432	678

Results from Table 8 show that as those are all popular websites; they are all getting large values in the inlinks (or also called backlinks). However, the large numbers in all countries such as (7320 and 7485) are for Google and YouTube which are popular websites in most countries and languages. Numbers in bold are for websites that are hosted in USA with values gathered from countries other than USA.

BJIS, Marília (SP), v.4, n.1, p.2-17, Jan./Jun. 2010. Available in: <http://www2.marilia.unesp.br/revistas/index.php/bjis/index>. ISSN: 1981-1640

With the exception of China, all other countries are getting around half of their popular websites from USA. There are a very high correlation between results collected from backlinks and the popularity of the website. However, outlinks did now show positive correlation in all cases with the website popularity.

4 CONCLUSION AND FUTURE WORK

This paper studied the effects of website location, and language on its popularity. The paper also evaluated the differences between the same search terms among different languages based on the country, and the language. English is still the de facto language in the Internet world. On the other hand, websites in the US, specially the popular ones, get universal popularity unlike websites in other countries.

Search engines are offering many services to other languages to allow equal opportunity for Internet users, no matter of their language or location. However, experiments and statistics gathered in this research showed that there are still some barriers for giving equal opportunities to websites regardless of their location, country, or language. On the other hand, many international websites have an English version for their website. Ultimately, search engines are expected to be designed in a way that make the language, or the location as pluggable features that can be changed at run time with the ability to translate all website content, images, icons, etc. to the new language dynamically.

In future, we will propose a new framework for search engine designs that takes into consideration language, and location plug-ability. A prototype search engine will be built and evaluated based on the proposed design.

REFERENCES

ALMAS, Y.; AHMAD, K. LoLo: A system based on terminology for multilingual information extraction. In: CALIFF, M. E. et al. **Coling Association of Computational Linguistics 2006**. In: WORKSHOP ON INFORMATION EXTRACTION BEYOND THE DOCUMENT, Sydney, Australia, 2006. Sydney: ACL, 2006. p.56-65

AL-ONAIZAN, Y.; KNIGHT, K. **Translating named entities using monolingual and bilingual resources**. In: PROCEEDINGS OF THE 40TH ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS (ACL), 40., Philadelphia, 2002. Philadelphia: ACL, 2002. p.400-408

BALFE, E.; SMYTH, B. A comparative analysis of query similarity metrics for community-based web search. In: CASE-BASED REASONING RESEARCH AND DEVELOPMENT, 3620, 2005. **Proceedings…** p.63-77

CHEW, P.; ABDELALI, A. **The effects of language relatedness on multilingual information retrieval**: A case study with Indo-European and Semitic languages. In: PROCEEDINGS OF THE WORKSHOP ON CROSS-LANGUAGE INFORMATION ACCESS, 2008.

CHRISTOF, M.; BONNIE, J.; DORR, M. **Iterative translation disambiguation for cross-language information retrieval**. In: PROCEEDINGS OF THE 28TH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 28., 2005. p.15-19

DANUSHKA, T. et al. **Measuring the similarity between implicit semantic relations from the web**. In: PROCEEDINGS OF THE 18TH INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 2009. Madrid, 2009.

DONG, Z. et al. A hybrid technique for English-Chinese cross language information retrieval. **ACM Transactions on Asian Language Information Processing (TALIP)**, v.7, n.2, p.1-35, 2008.

ERINJERI, J. P. et al. Development of a Google-based search engine for data mining radiology reports. **Journal Digit Imaging**, v.22, p.348-356, Apr. 2008.

GUO, W.; BIAN, S.; YUAN.T. Integrating query translation and text classification in a cross-language patent access system. In: PROCEEDINGS OF NTCIR-7 WORKSHOP MEETING, 2008. p.16-19

JACQUES, S. Comparative study of monolingual and multilingual search models for use with Asian languages. **ACM Transactions on Asian Language Information Processing (TALIP)**, v.4, n.2, p.163-189, 2005.

JANG, M. G. et al. **Simple query translation methods for Korean-English and Korean-Chinese CLIR in NTCIR experiments**. In: WORKING NOTES OF THE THIRD NTCIR WORKSHOP MEETING – PARTII: CROSS. 2002.

LIANHAU, L. et al. **Mars:** Multilingual access and retrieval system with enhanced query translation and document retrieval. In: THE 47TH ANNUAL MEETING OF ACL

AND THE 4TH INTERNATIONAL JOINT CONFERENCE OF NLP (SW DEMO), 47., Singapore. Singapore: 2009. p.21-24

LIN, F. et al. **Query formulation with a search assistant**. In: ICADL, LNCS 3334. 2004. p.491-500

LOIA, V.; SENATORE, S. Customized query response for an improved web search. In: CASTILLO, O. (Ed.). **Theory advance and applied of fuzzy logic**. ASC 42, 2007. p.653-662

NILSSON, K.; HJELM, H.; OXHAMMAR, H. **Cross-language ontology-driven information retrieval in a restricted domain**. In: PROCEEDINGS OF THE 15TH NODALIDA CONFERENCE, 15., 2005. p.139-145

SALTON, G. **Automatic processing of foreign language documents**. In: PROCEEDINGS OF THE 1969 CONFERENCE ON COMPUTATIONAL LINGUISTICS. INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS. ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. Morristown, 1969 p.1-28

ZHO, W.; YU, C.; MENG W. **A system for finding biological entities that satisfy certain conditions from texts**. In: CIKM'08. Napa Valley (CA), 2008. p.1281-1290

Brazilian Journal of Information Science

Anas AlSobh
Department of Computer Information Systems (CIS)
Faculty of Information Technology and Computer Sciences
Yarmouk University
Jordan
E-mail:

Ahmed Al Oroud
Department of Computer Information Systems (CIS)
Faculty of Information Technology and Computer Sciences
Yarmouk University
Jordan
E-mail:

Mohammed N. Al-Kabi
Department of Computer Information Systems (CIS)
Faculty of Information Technology and Computer Sciences
Yarmouk University
Jordan
E-mail:

Izzat AlSmadi
Department of Computer Information Systems (CIS)
Faculty of Information Technology and Computer Sciences
Yarmouk University
Jordan
E-mail: alsmadi@gmail.com