IASET: International Journal of Library & Educational Science (IASET: IJLES)
ISSN(P): Applied; ISSN(E): Applied
Vol. 2, Issue 2, Jul – Dec 2016, 61-70

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

CURRENCY OF RESEARCH ARTICLES FOR SELECT MAJOR SEARCH ENGINES IN THE FIELD OF LIBRARY & INFORMATION SCIENCE

PEERZADA MOHAMMAD IQBAL[1] & ABDUL MAJID BABA[2]

[1]Professional Assistant, Faculty of Fisheries Library, Sher-e-Kashmir University of Agricultural Sciences & Technology of Kashmir (SKUAST-K), India

[2]Head, Department of Library and Information Science, The University of Kashmir, India

ABSTRACT

This paper presents the results of a research conducted on seven search engines- Google, Bing, Yahoo, Ask, Baidu, Dogpile and DuckDuckgo for the currency of scholarly articles using Library and Information Science related search terms. The search engines are evaluated by taking the first twenty results pertaining to 'scholarly articles' for estimation of currency of literature. It shows that 'Baidu' performance is better in retrieving current research documents while 'Google' retrieves highest number of dated scholarly documents.

KEYWORDS: Currency, Date, Library & Information Science, Search Engine, Modification, Update

INTRODUCTION

The World Wide Web can be used as a quick and direct reference to get any type of information in electronic format all over the world. However, information found on the Web needs to be filtered and may include voluminous misinformation or non relevant information. The user or Internet surfer may not be aware of quality search engines to get information on a topic quickly and may use different search strategies. Finding useful information quickly on the Internet poses a challenge to both the ordinary users and the information professionals. Though the performance of currently available search engines has been improving continuously with powerful search capabilities of various types, the lack of updateness, the inability to predict the quality of retrieved results, and the absence of controlled vocabularies make it difficult for users to use search engines effectively. The use of the Internet as an information resource needs to be carefully evaluated as no traditional quality standards or control have been applied to the Web. Librarians need to be able to provide informative recommendations to their clientele regarding the selection of search engines and their effective search strategies. A main hand in use for librarians is Current Awareness Service (CAS), an information service that is updated as per routine. For this service librarians mainly depend on search engines to give the said service.

Problem

In the beginning of the internet, it was easy to fine information using variety of software that was usually command driven rather than using a graphical interface. With the proliferation of information, systems such as Archie, Gopher and Veronica became increasingly unable to cope with huge information. The advent of many types of search engines provided solution for literature search using Boolean operators, Proximity searching, Wild cards, Truncation etc. Many search engines developed new versions and techniques to achieve some kind of sophistication but all have not helped to forward the case of access and searching from scholar's perspective. Besides keeping in view different ways of

indexing the internet, search engines operate in different ways and retrieve documents in different orders. Further, it does not sift information form scholar's point of view i.e., it retrieves information on a particular topic from different aspects like marketing, advertisement, news and entertainment mixed with some research papers. The academic community attempts to look purely for scholarly information on his topic of interest to have output/ retrieval best in terms of Currency.

The present investigation attempts to evaluate the performance of the select search engines in retrieving scholarly research articles in the field of Library and Information science with respect to their currency of research articles.

OBJECTIVES

The following objectives are laid down for the study:

- **To Select Search Engines & Search Terms for the Study**

There are countless numbers of search engines over the internet. Some are active while others are inactive, some are country bound other are global, some are subjective, unilingual, etc while others are general, multilingual etc. Selection of search engines will be based on the following parameters.

- Automatic Indexing.
- Global Coverage.
- Advanced search feature.
- Refine searching in Portable Document Format (PDF).
- Providing gist of information while indexing

Since the scope of the study relates to the field of Library and Information Science. The terms will be selected using classifying schemes from Library and Information Science and List of subject headings. The terms will be further refined to into three categories i.e., Simple, Compound and Complex terms.

- **To Find Out the Currency of "Scholarly Research Articles" Retrievable Through Selected Search Engines**

The study will estimate total results and scholarly documents retrieved from provided search terms with currency of the engines by analyzing the dates pertaining to publication/ modification of the retrieved scholarly publication.

Methodology

As certified by International Standard Organization there are 230 search engines **(Promote3.com, 2015)** available for searching the web. These search engines are of various types like general search engine, robotic search engine, Meta search engine, directories and specialized search engines. Most users prefer robotic search engines as they allow the users to compose their own quires rather than simply follow pre specified search paths or hierarchy as in case of directories. Moreover, robotic search engines locate data in a similar way i.e., by the use of crawlers or worms. This distinguishing feature differentiates them form web directories like Yahoo! Where collections of links to retrieve URL's are created and maintained by subject experts or by means of some automated indexing process. However some of these services are also include a robot driven search engine facility. But this is not their primary purposes. This due to this feature Yahoo! Was included for the study.

Meta search engine e.g., Dogpile etc don't have their own database. These access the database of many robotic search engines simultaneously. Thus these were included for the study.

Still hundreds of robotic general search engines navigate the web, in order to limit the scope of study after preliminary study, following criteria was laid down for selection of general search engines:

- Availability of automated indexing
- Global coverage to data.
- Quick response time.
- Availability of filter search mechanism
- Least overlapping.
- Major market holder.

Following two general search engines were selected for the study for meeting all the criteria and being comprehensive in nature.

a) Google. b) Baidu.

Since the study relates to the field of Library and Information Science. It was felt to include specialized search engine in the study representing question answer search engine i.e., Ask.com & another specific i.e., Bing. There being no full-fledged search engine in the field Library and Information Science except many associated with library websites. Among those human powered (DuckDuckGo) after preliminary investigation and feasibility in the study was included in the study. Thus the search engines undertaken for evaluation of study are:-

- Google (General)
- Bing (Specific)
- Yahoo!(Directory)
- Ask (Question Answer Search engine)
- Baidu (Country Specific General Search engine)
- Dogpile (Meta search Engine)
- DuckDuckgo (Human Powered Search Engine).

Selection of Terms

Selection of terms is not directly possible in development and multidimensional field like Library and Information Science. Therefore, classification schemes like DDC (18^{th}) and DDC (22^{nd}) were consulted to understand Broad/Narrow structure of Library and Information Science. It helped to get five terms/Fields i.e.,

- Information System.
- Digital Library.

- Library Automation.
- Library Services.
- Librarianship.

These terms were then browsed in "LC list of subject Headings" which provided many other related terms (RT) and Narrow terms (NT). Further NT and RT attached to each other preferred or standard terms were also browsed which retrieve a large number of Library and Information Science terms. At first instance 140 Library and Information Science related terms were identified.

Some terms occurred more than once and duplication removed. It reduced the number to 100. Later terms were divided into three broad groups under:

- Application.
- Transformation.
- Inter-relation.

"Application" denotes utility of Library and Information science in various fields and about 50 terms came under this group. "Transformation" refers to a method of developing or manufacturing library services into practical market and 30 terms fall under this group. "Inter-relation" means transformation/dependence of one subject onto another and 20 terms came under this group. Further each category is sub-divided into groups.

"Application" into four i.e., "Reference service", "Informatics", "Information Retrieval" & "Information Sources". "Transformation" into two i.e., "Digitization" & "Consortia". "Inter-relation" into two i.e., "Library Network" & "Information System".

The terms in each group were arranged alphabetically and each term was given a tag. Later 20% of the terms were selected from each group using "Systematic Sampling" (i.e., first item selected randomly and next item after specific intervals). It further reduced the number to 19. Finally the selected terms were classified into three groups under "Simple", "Compound" & "Complex Terms" (Table 1.1). This was done in order to investigate how search engines control and handle simple and phrased terms. "Simple Terms" containing a single word were submitted to the search engine in the natural form i.e., without punctuating marks. "Compound Terms" consisting of two words were submitted to the search engines in the form of phrases as suggested by respective search engines and "Complex Terms" composed of more than two words or phrases, were sent to the search engine with suitable Boolean operator "AND" & "OR" between the terms to perform special searches.

Table 1.1: Keywords

S. No	Simple Terms	Compound Terms	Complex Terms
1	Catchwork	Bibliometric Classification	Digital Library Open Source Software
2	Citation	Citation Analysis	Health Information System
3	Dublincore	Comparative Librarianship	Library Information System
4	Indexing	Digital Preservation	Library Information Network
5	Manuscript	Electronic Repositories	Multimedia Information Retrieval
6	Plagiarism	Library Automation	
7	Reprints	Semantic web	

Selection of Search Results and Filtration Technique

To evaluate the select search engines top 20 results from each search engine was taken into consideration to determine precision. The assessment of top 20 results is supported by **Hawking, Craswell, Bailey & Griffiths, (2001)** compared 20 search engines using first top 20 search results comparing 54 topics originated by anonymous searchers for measuring search engine qualities. **Tongchim, Sornlertlamvanich & Isahara, (2006)** used seven search engines for measuring effectiveness of search engines on Thai queries. Their results calculated from binary relevance judgments of the first 20 returned results, using 56 topics. Latter **Egelman, S., Cranor, L., & Chowdhury, A. (2006)** conducted a study of quality and quantity of (Platform for privacy preferred project) P3P-encoded policies associated with top-20 search results from three popular search engines viz., AOL, Google, and Yahoo!. The study examined top 20 search results returned by each search engine to build a P3P-enabled search engine and used it to gather statistics on P3P adoption as well as the privacy landscape of the Internet as a whole. **Dirk (2008)** Evaluated 5 search engines using first top 20 hits for retrieval effectiveness of web search engines.

Andago, Phoebe & Thanoun, (2010) collected queries from 30 university students and entered these queries into two search engines viz., Google and Hakia. Precision was thereafter calculated using first 20 hits. The 20 results were taken into evaluated for a comparison of precision of Semantic Search Engine against a Keyword Search Engine. **Ajayi & Elegbeleye (2014)** used first 20 results for performance evaluation of three search engines. The use of first 20 results were thought to be genuine as majority of scholars use first two pages of search hits which by default to many search engines is fixed to 10 hits per page. The evaluation of first top 20 hits was further backed by a questioner among the scholars of Kashmir University. A total of 100 questioners were distributed among the doctorial and M.Phil scholar of said university. The aim was to check the result extension at maximum and type of filtration a scholar uses. It was revealed that 84 percent of the scholar prefer first 20 hits (or two pages: a default of 10 results per page), 11 percent prefer first 10 results and five percent prefer more than 20 results. Further it was revealed that scholar's use PDF (portable document format) to filter the results as to get maximum of research article.

Currency (Publication/ Modification Date of Research Articles)

The web is an ocean of information which may become outdate in no time. The currency of a document is defined as a measure of reviewing a search engine index to incorporate current documents. This is significant for all search engines in order to provide latest information. **Table 1.0** demonstrates the currency of search results of the search engines linked for scholarly documents. The date considered is either the date of publication of a document or the date on which a document is modified. However, the later was preferred where it was available.

Table 1.0: Currency of Publication

Total	Google	Bing	Yahoo!	Ask	Baidu	Dogpile	DuckDuckGo
2015	5 (17.86 %)	3 (10.71 %)	3 (10.71 %)	3 (10.71 %)	9 (32.14 %)	3 (10.71 %)	2 (7.14 %)
2014	5 (15.63 %)	4 (12.50 %)	3 (9.38 %)	9 (28.13 %)	4 (12.50 %)	3 (9.38 %)	4 (12.50 %)
2013	4 (8.70 %)	6 (13.04 %)	8 (17.39 %)	6 (13.04 %)	8 (17.39 %)	6 (13.04 %)	8 (17.39 %)
2012	5 (10.00 %)	8 (16.00 %)	6 (12.00 %)	7 (14.00 %)	9 (18.00 %)	8 (16.00 %)	7 (14.00 %)
2011	2 (6.90 %)	4 (13.79 %)	5 (17.24 %)	3 (10.34 %)	4 (13.79 %)	7 (24.14 %)	4 (13.79 %)
2010	5 (14.71 %)	4 (11.76 %)	7 (20.59 %)	5 (14.71 %)	4 (11.76 %)	5 (14.71 %)	4 (11.76 %)
Below 2010	54 (15.93 %)	46 (13.57 %)	40 (11.80 %)	56 (16.52 %)	67 (19.76 %)	39 (11.50 %)	37 (10.91 %)
Undated	5 (8.06 %)	8 (12.90 %)	9 (14.52 %)	4 (6.45 %)	13 (20.97 %)	16 (25.81 %)	7 (11.29 %)
Total	85 (13.71 %)	83(13.39 %)	81(13.06 %)	93(15.00 %)	118(19.03 %)	87(14.03 %)	73(11.77 %)

Figures in Parenthesis Indicate Percentage

It is evident from the table that Baidu provides greatest freshness with 44.64% of the documents published or modified in between 2014-2015, 35.39% published or modified in between 2012-2013, 25.56% appearing during 2010-2011 and 19.76% before 2010 while as 29.97% of the documents do not provide the date of publication or modification.

Ask shows 38.84% of the retrieved documents published in between 2014-2015, 27.04% in between 2012-2013, 25.05% in 2010-2011 and 16.52% of the documents published before 2010, while only 6.45% of the scholarly documents do not exhibit any date of publication or modification.

Google Retrieved 33.48% of the scholarly documents published or modified in between 2014-2015, 18.70% of scholarly documents in between 2012-2013, 21.60% in between 2010-2011 and 15.93% of the documents published before 2010, while only 8.06% of the scholarly articles do not have any date of publication or modification.

Bing shows 23.21% of retrieved scholarly publication published or modified in between 2014-2015, 29.04 of documents in between 2012-2013, 25.56% in between 2010-2011 and 13.57 of documents published before 2010, while only 12.90% of the articles don not have any date of publication or modification.

Yahoo! And Dogpile remarkably retrieved similar percentage of scholarly documents published or modified. 20.09 % of documents from 2014-2015, 29.39% and 29.04% in between 2012-2013, 37.83% and 38.84% in between 2010-2011 and 11.80% and 11.50% below 2010 while a difference lies in the undated portion. Yahoo! Retrieved 14.52%, while as a major portion of the scholarly articles 25.81% of Ask search engine were undated.

DuckDuckGo retrieved 19.64% of the scholarly documents published or modified in between 2014-2015, 31.39% of scholarly documents in between 2012-2013, 25.56% in between 2010-2011 and 10.91% of the documents published before 2010, while only 11.29% of the scholarly articles do not have any date of publication or modification.

While comparing the overall currency of all 7 search engines (i.e., documents published or modified in between 2010-2015). DuckDuckGo shows the highest currency (39.73%) followed by Yahoo! (39.51%) and Dogpile (36.78%). Ask has currency of (35.48%) while Bing has (34.94%). Baidu show (32.20%) of currency of documents, and shockingly Google show the least currency of (30.59%).

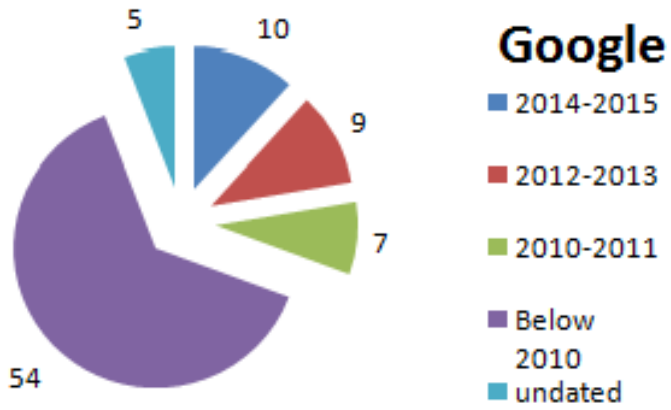

Figure 2.0: Currency of Google Search Engine

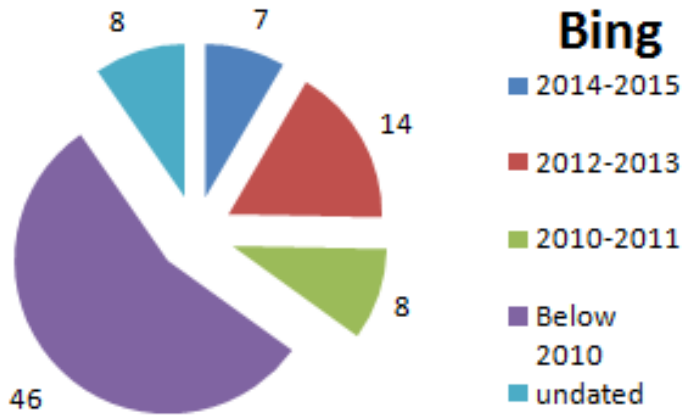

Figure 3.0: Currency of Bing Search Engine

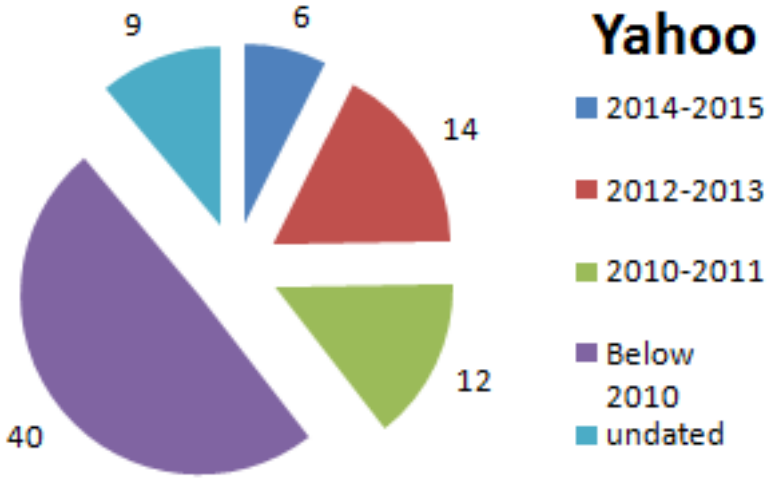

Figure 4.0: Currency of Yahoo! Search Engine

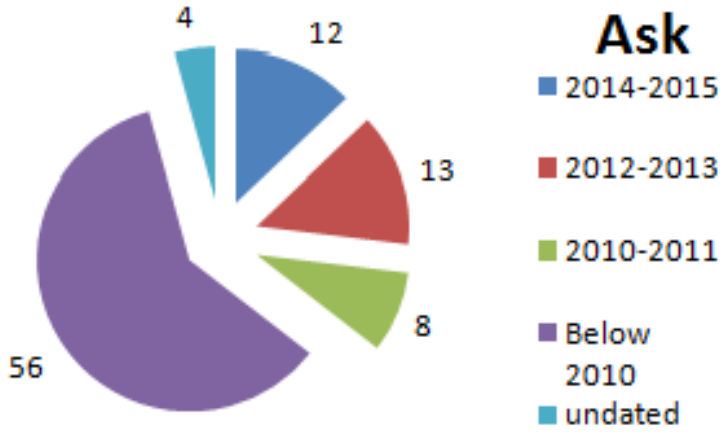

Figure 5.0: Currency of Ask Search Engine

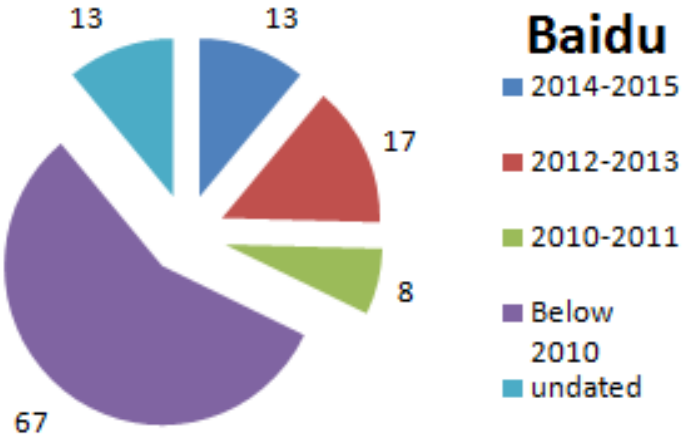

Figure 6.0: Currency of Baidu Search Engine

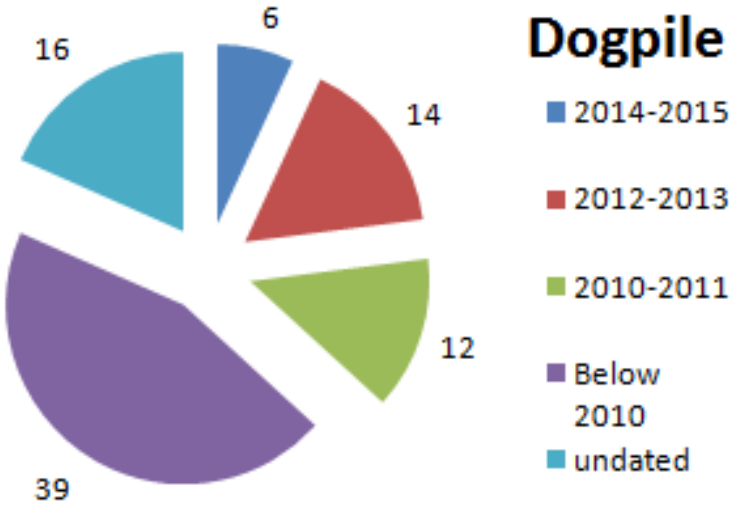

Figure 7.0: Currency of Dogpile Search Engine

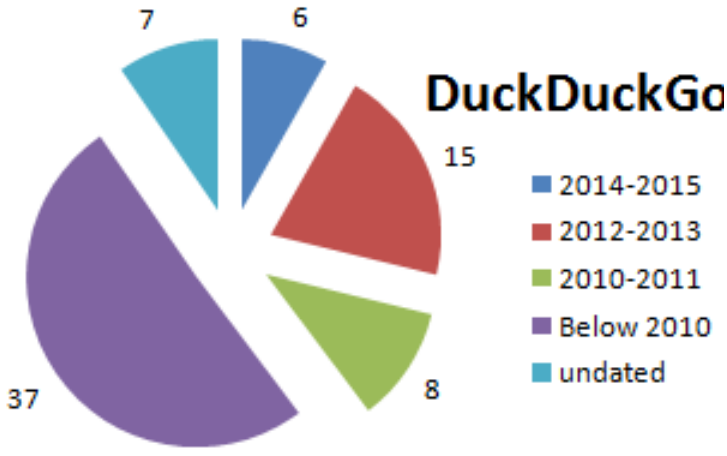

Figure 8.0: Currency of Duckduckgo Search Engine

FINDINGS AND CONCLUSIONS

The study traces the performance of the select search engines in retrieving currency of scholarly publications in the field of Library and Information Science. It has evaluated the engines with respect to their publication date or modification. The findings based on observation, experimentation and statistical analysis of data are enumerated as:

- Baidu performance is better in retrieving current research documents followed by Ask and Dogpile. Google, Bing and Yahoo show similar currency while DuckDuckGo is the least current.
- Undated information is no way better than anonymous information. Among the general search engine Google retrieves highest number of (94.12%) dated scholarly documents followed by Bing (90.36%) and Baidu (88.98%). The highest dated scholarly information among all select search engines is retrieved by Ask (95.70%) whereas the highest number of undated information is retrieved by Dogpile (18.02%) and Yahoo! (11.11%) respectively. The human powered DuckDuckGo search engine also showed a good currency (90.41%).

REFERENCES

1. Ajayi, O. O., & Elegbeleye, D. M. (2014). Performance Evaluation of Selected Search Engines. *Computer Engineering and Intelligent Systems*, 5(1), Retrieved from

   http://www.iiste.org/Journals/index.php/CEIS/article/viewFile/10301/10504

2. Andago, M.O., Phoebe, T., & Thanoun, B.A.M. (2010). Evaluation of a Semantic Search Engine against a Keyword Search Engine Using First 20 Precision. *International Journal for the Advancement of Science & Arts*, 1(2), 55-63. Retrieved from http://www.ucsiuniversity.edu.my/cervie/pdf/ijasa/paperV1N2IT3.pdf

3. Dirk, L. (2008). The Retrieval Effectiveness of Web Search Engines: Considering Results Descriptions. *Journal of Documentation*, 64(6), 915 – 937. DOI: 10.1108/00220410810912451

4. Egelman, S., Cranor, L., & Chowdhury, A. (2006). An Analysis of P3P-Enabled Web Sites among Top-20 Search Results. In *proceedings of the 8th International Conference on Electronic Commerce* (pp. 197 - 207). Retrieved from http://casos.cs.cmu.edu/publications/papers/icec06.pdf

5. Hawking, D., Craswell, N., Bailey, P., & Griffiths, K. (2001). Measuring Search Engine Quality. *Information Retrieval*, 4(1), 33–59. DOI: 10.1023/A:1011468107287

6. Promote3.com (2015). *Top Search Engine Ranking Search Engine Optimization*. IDV International: California. Retrieved from http://www.promote3.com/search-engine-230.htm

7. Tongchim, S., Sornlertlamvanich, V., & Isahara, H. (2006). Measuring the Effectiveness of Public Search Engines on Thai Queries. In: *Proceedings of The Fifth IASTED International Conference on communications, internet, and information technology*. Retrieved from

   http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.1951&rep=rep1&type=pdf