

MALAYSIAN WEB SEARCH ENGINES: A CRITICAL ANALYSIS

Hananzita Halim and Kiran Kaur

MLIS Programme, Faculty of Computer Science & Information Technology,
University of Malaya, Kuala Lumpur, Malaysia
e-mail: rashan@gmail.com, kiran@um.edu.my

ABSTRACT

This paper reports the results of a study conducted to explore and compare the features of independently built Malaysian Web search engines, as well as evaluate their performance and search capabilities. Four Malaysian independently built search engines were chosen for the study and Google served as the benchmark engine. This study applied two co-related methodologies, namely Information Retrieval Evaluation Methodology and common features comparison. The Information Retrieval Evaluation Methodology measures include search capabilities, retrieval performance (recall & precision) and user effort. Results show that each search engine portrays different strengths and weaknesses. Malaysia Directory appears to be the best Malaysian Web search engine in term of feature presentation but fails in term of retrieval performance; whereas Malaysia Central managed to score the highest for mean precision but failed to excel in terms of feature presentation. SajaSearch outperformed Cari in most related documents, even though Cari did return the highest total hits. Overall comparison with Google leaves Malaysian search engines with much to improve on in terms of performance and search capabilities. Implications for the design of Malaysian Web Search Engines are also discussed.

Keywords: Web search engine; Information retrieval; Website evaluation; Malaysia

INTRODUCTION

The explosive growth of the Internet has rendered the World Wide Web as the primary tool for information retrieval today. The World Wide Web is revolutionizing the way people access information, and has opened up new possibilities in areas such as digital libraries, information dissemination and retrieval, education, commerce, entertainment, government, and health care (Lawrence and Giles 1999). Various Web search aids have been developed in order

Hananzita, H. & Kiran K.

to provide users with an interface that enables them to locate documents containing information that matches their interests. Although general popular Web search engines, such as Google, Yahoo! and AltaVista are getting easier to use, sometimes locating relevant information especially local information is still like looking for a needle in a haystack. A growing number of countries are beginning to develop their own search aids to facilitate the search for local content. These varieties of search aids consist of search engines with subject directories, search directories and specialized search engines with subject directories. The efficient retrieval of related information is a major research goal of library and information science. Evaluating these Web search engines becomes increasingly important especially if it helps to answer questions concerning both the way they work and the accuracy of the results they provide. Though many comparative search engine studies have been conducted and published, little has been done on Malaysian Web search engines. This study focuses on one of the information retrieval (IR) basic areas of research, that is evaluation. The researchers intend to explore, compare and evaluate the currently available local Web search engines in Malaysia. This study also aims to exploit more efficiently the features and capabilities of those engines available for IR.

According to Rijsbergen (1979), there are three main areas of research in IR: content analysis; information structures; and evaluation. Briefly, the first is concerned with describing the contents of documents in a form suitable for computer processing; the second with exploiting relationships between documents to improve the efficiency and effectiveness of retrieval strategies; the third with the measurement of the effectiveness of retrieval. Evaluation is one of the three basic areas of research in IR. Evaluation of IR is mainly concerned with the measurement of effectiveness of retrieval. Much of the research and development in information retrieval is aimed at improving the effectiveness and efficiency of retrieval. Thus, much effort and research has gone into solving the problem of evaluation of information retrieval systems. In order to evaluate an IR system, it is important to know what can be measured that will reflect the ability of the system to satisfy the user. According to Rijsbergen (1979), as early as 1966 Cleverdon listed six main measurable quantities:

- (a) the coverage of the collection, that is, the extent to which the system includes relevant matter;
- (b) the time lag, that is, the average interval between the time the search request is made and the time an answer is given;
- (c) the form of presentation of the output;
- (d) the effort involved on the part of the user in obtaining answers to his search requests;

- (e) the recall of the system, that is the proportion of relevant material actually retrieved in answer to a search request;
- (f) the precision of the system, that is, the proportion of retrieved material that is actually relevant.

Of these criteria, recall and precision have most frequently been applied in measuring IR (Gwizdka and Chignell, 1999). Consistent to Rijsbergen's statement, Wang (2001) also stresses that a very good quality search engine is said to have high precision and recall. His definition of precision is a measure of the usefulness of a hit list while recall is a measure of the completeness of the hit list. Recall is a measure of how well the engine performs in finding relevant documents. Recall is 100% when every relevant document is retrieved. In theory, it is easy to achieve good recall: simply return every document in the collection for every query. Therefore, recall by itself is not a good measure of the quality of a search engine. Precision is a measure of how well the engine performs in not returning non-relevant documents. The relationship between precision and recall is described in Figure 1.

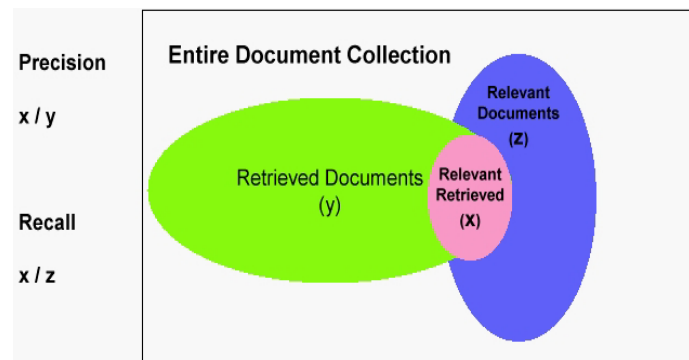


Figure 1: Relationship between Precision and Recall

Search Engine Evaluation Efforts

Even though several local Web search engines in Malaysia have emerged and become popular for the past eight years, unfortunately, there is little published comparative and evaluative study on these engines. Since the first public Web search engine existed in 1994, there have been a tremendous number of comparative and evaluation studies done on them by researchers all around the world. Yet, to date, all the comparative and evaluation studies conducted and published mainly involve the already well-known Web search engines such as Google, AltaVista,

Hananzita, H. & Kiran K.

InfoSeek and Excite. There are even studies that compared the search features and evaluated the retrieval performance of Web search engines designed for children.

Davis (1996) provides an extensive review on comparison of seven search engines, AltaVista, Hotbot, Infoseek, Excite, Lycos, Open Text and WebCrawler. The comparisons were based on the search engines features and characteristics only. The study reported by Botluk (2000) compared six major Web search engines namely AltaVista, Excite, Go, Google, HotBot and Lycos. The author's focus was, similar to Eric's, on describing the features of these various search engines. She described each engine in detail as well as produced a search engine comparison chart. The evaluation criteria concerned in this study were search language, search restrictors, results display, subject directory, and other search features.

There has been notable research of web search engines based on precision as a criteria for evaluation. Leighton (1996) did a study of Web search engines for course work, actually employing the evaluation criterion of precision. He evaluated Infoseek, Lycos, WebCrawler and WWWorm using 8 reference questions from a university library as search queries. In a later study following the earlier study, Leighton (1997) conducted another project with the same objective focusing on precision of Web search engines. This time around, he compared the precision of five commercial Web search engines namely AltaVista, Excite, HotBot, Infoseek and Lycos. The measurement used, "first twenty precision" and rated the engines based on the percentage of results within the first twenty returned that were relevant or useful. Chu and Rosenthal (1996) followed Leighton's footsteps in evaluating Web search engines by applying precision as one of the evaluation criteria. In their comparative study, they compared and evaluated three Web search engines, namely Alta Vista, Excite and Lycos. The comparison and evaluation were in terms of their search capabilities and retrieval performance. One evaluation criterion of IR, recall, was deliberately omitted from this study because it was impossible to assume how many relevant items there are for a particular query in the huge and ever changing Web system. At the end of the study, the authors reasoned out which was the best engine, which was Alta Vista, as well as proposed a methodology for evaluating other Web search engines.

Later Wishard (1998) made the most impressive comparative study using precision when she compared thirty-seven (37) Internet search engines. The search suite used in the study is small but it is specific in its focus on earth science related subjects only. The author included an evaluation of search engines' precision based upon three sample searches, but did not include the exact precision figures, in which the

precision was indicated as high, average or low. It was concluded that no one search engine emerges as the most precise for locating information on the World Wide Web. This conclusion concurs with other precision studies. No one tool emerged preeminent, even though the author has focused on only the earth science queries. Hawking et al (1999) also performed a comparative study using precision in finding out which search engine is best at finding online services. Eleven search engines including Google, Fast, NorthernLight, Lycos and AltaVista were compared. The results lists for each engine were evaluated using precision of ten documents retrieved. This study is known as the first published study to investigate search engine performance on online service queries. Sutachun (2000) in his thesis compared five search engines: AltaVista, Excite, HotBot, Lycos and Infoseek in terms of search features and retrieval performance. He also applied the precision method in evaluating the effectiveness of each engine. Shang and Li (2002) conducted an explicit study on search engines evaluation by focusing precision of each engine. The authors evaluated the search engines in two steps based on a large number of sample queries: (a) computing relevance scores of hits from each search engine; and (b) ranking the search engine based on statistical comparison of the relevance scores. Six popular search engines were evaluated based on queries from two domains of interest, parallel and distributed processing and knowledge and data engineering. The results showed that overall Google was the best.

Apart from precision methods, there are also studies that only concentrated on evaluating search engines performances by using the user effort evaluation measurement. Tang and Sun (2000) applied the three user-effort-sensitive evaluation measures on four Web search engines. The three user-effort measures are “first twenty full precision”, “search length”, and “rank correlation”. The authors argued that these measures are better alternatives than precision and recall in Web search situations because of their emphasis on the quality of ranking. Besides using the conventional evaluation criteria methods, research shows that starting in 1999, more and more new evaluation criteria methods were being proposed and introduced in order to evaluate the performance of search engines. Losee and Paris (1999) proposed that instead of using traditional performance measures such as precision and recall, IR performance may be measured by considering the probability that the search engine is optimal and the difficulty associated with retrieving documents with a given query or on a given topic.

The study reported in Hashim and Yusof (2000) evaluates ten search engines using precision. Besides the precision measurement, the authors also introduced an overlap measurement to determine the commonality of documents between the hit

Hananzita, H. & Kiran K.

lists of various search engines. They reported the correlation between the ranking list and the overlap measurement. The report showed that the top five ranked search engines based on precision measurement having the most number of documents in common. Chowdhury and Soboroff (2002) presented a method for comparing search engines automatically based on how they rank known item search results. The method uses known item searching; comparing the relative ranks of the items in the search engine's rankings. The approach automatically constructs known item queries using query log analysis and automatically constructs the result via analysis editor comments from the Open Directory Project. Five well-known search services were compared in this study namely Lycos, Netscape, Fast, Google, and HotBot.

In a more recent publication, Ohtsuka, Eguchi and Yaware (2004) proposed "a user oriented evaluation criterion" that evaluates the performance of Web search systems by considering users actions when they retrieve Web pages. The authors also went a step further by evaluating the proposed criterion in comparison with the conventional methods by measuring the time spent on search as the users' satisfaction degree. Similarly, Sugiyama, Hatano and Yoshikawa (2004) also proposed evaluation methods based on user's need. The authors proposed several approaches to adapting search results according to each user's need for relevant information without any user effort. The approaches involved in their experiments were (a) relevance feedback and implicit approaches, (b) user profiles based on pure browsing history, and (c) user profiles based on the modified collaborative filtering. Their approaches allow each user to perform a fine-grained search, by capturing changes in each user's preferences.

Even though numerous comparisons and evaluations were performed on search engines, there is no well-defined and standard methodology for those studies. Moreover, most of the methods of evaluation are frequently not fully specified in the published reports. However, from the comparative and evaluation studies discussed above, it can be inferred that there are apparently two types of evaluation studies of Web search engines: characteristics evaluation and performance evaluation. Thus, methodologies for this study were derived from this fact.

Internet and Web Search Engines in Malaysia

Malaysia is clearly ahead of many countries in terms of personal computer ownerships, Internet usage and in value usage of Internet. Malaysia currently has an estimated 10 million Internet users, that's about 41% of its total population (Internet World Stat, 2005). Over the last five years, Internet users have increased from 15% to 41% of the total population. In 2004, Malaysia was at the eighteenth position in

Malaysian Web Search Engines

the top twenty (20) countries in the world with the highest number of Internet users (Internet World Stat, 2004). The phenomenal growth in the size of Internet users has created a notable number of Malaysian Web search engines and directories. Malaysian Web search engines refer to search engines that are specially targeted or focus on Malaysia related only web sites and homepages. From the list compiled, it may be concluded that Malaysian Web search services can be divided into two categories namely custom-made search engines or directories and independent search engines or directories (Table 1).

Custom-made search engines or directories refer to engines or directories that use other established search engines such as Yahoo!, Google or AltaVista to produce search results. In other words, based on these engines, the local engines are customized accordingly just to perform search on Malaysian website and homepages only. Meanwhile, independent search engines or directories are those engines or directories that are independently, locally built to specially locate Malaysian Web sites and homepages. This type of engines does not rely on those other established and popular engines to produce results.

Table 1: Types of Search Services in Malaysia

| Search Services | Custom made | Independent |
|---------------------------------|---|--|
| Search Engine | Google Malaysia | <i>Cari Malaysia Directory Malaysia Central SajaSearch</i> |
| Search Directory | Catcha Gotcha Skali Malaysia Malaysia Focus | Mesra Centre U2asean Malaysia eGuide |
| Search Engines With Directories | Yahoo!Malaysia New Malaysia Asiaco Malaysia WebPortal AsiaDragons Malaysia Everything Malaysia | AsiaNet Malaysia |

Although the number of Malaysian search engines and directories show an impressive figure, this study focused only on those independent locally built search engines, eliminating those dependent and independent directories. The researchers would like to focus only on search engines that should at least allow users to compose their own search queries rather than simply follow pre-specified search

Hananzita, H. & Kiran K.

paths or hierarchy as in the case of those independently built search directories. Thus, only four (4) Malaysian Web search engines were chosen for this study namely *Cari*, *Malaysia Directory*, *Malaysia Central* and *SajaSearch*.

This study is evidently significant in order to provide practical assistance for search engine users, especially for our local users. It is not intended to serve for the purpose of Malaysian users only but also for the benefits of local search engines developers. Hopefully, this study can act as a reference tool in order to design better and improved search engines that can outperform those well-known and frequently used search engines.

THE STUDY

The comparative study of the Malaysian Web search engines was conducted with the attempt to investigate and evaluate each local Web search engine features and capabilities; and to compare and contrast each Web search engine in order to provide users (especially other researchers) a platform for further investigation on effectiveness of Web search engines. Specifically, the study aimed to answer the following research questions:

- a) What features do each of the Malaysian Web Search Engines offer?
- b) What are the functionalities and capabilities of Malaysian Web Search Engines in terms of recall and precision?
- c) What are the strengths and weaknesses of Malaysian Web Search Engine?

This study used the descriptive research method in which the researchers systematically described the background, features, functionalities and capabilities of each selected Malaysian Web Search Engines. For evaluation and analysis purposes, two (2) co-related methods namely Information Retrieval Evaluation Methodology and common features comparison were applied. This IR Evaluation Methodology comprises of three techniques on how to evaluate search engines; namely ‘search capability’, ‘retrieval performance’ as well as ‘user effort’. While the features or criteria that are compared among the engines are system description, results display, subject directory, and other special features. The research was limited only on search engines that should at least allow users to compose their own search queries rather than simply follow pre-specified search paths or hierarchy as in the case of those independently built search directories. Thus, only four (4) Malaysian Web search engines were chosen for this study namely *Cari*, *Malaysia Directory*, *Malaysia*

Central and *SajaSearch*. All four search engines were evaluated using Google as a bench mark.

RESULTS AND DISCUSSION

Search Engines Features

Search engine features and capabilities are compared to Google using three main categories of criteria, “search capability”, “result display” and “user effort”. Table 2 illustrates the total scores attained by each of the search engines compared. Y represents the existence of the feature and N represents the non-existence of the feature. For each Y answer a mark of 1 is given while 0 is given for each N answer.

Table 2: Features Comparison Score

| Evaluation Criteria | Cari | MyCen | MyDir | Saja | Google |
|-----------------------------|----------|----------|----------|----------|-----------|
| Search Capability | | | | | |
| a. Boolean | Y(1) | N | Y(1) | N | Y(1) |
| b. Truncation/Wildcard | N | Y(1) | N | N | Y(1) |
| c. Field Search | N | N | N | N | |
| d. Keyword or Phrase Search | N | N | Y(1) | N | Y(1) |
| e. Search Restrictor/Limit | Y(1) | N | Y(1) | N | Y(1) |
| Results Display | | | | | |
| a. Short Summary | Y(1) | Y(1) | Y(1) | Y(1) | Y(1) |
| b. URL | N | Y(1) | N | Y(1) | Y(1) |
| c. Size | N | N | N | N | Y(1) |
| d. Page date | N | N | N | N | Y(1) |
| e. File type | N | N | N | Y(1) | Y(1) |
| User Effort | | | | | |
| a. User Aids | N | Y(1) | Y(1) | N | Y(1) |
| b. Subject Directory | Y(1) | Y(1) | Y(1) | N | N |
| Total Score | 4 | 5 | 6 | 3 | 10 |

Malaysia Directory outperforms other search engine in terms of features presentation. *Malaysia Directory* offers more search capabilities than others, which can be found in its advanced search feature. *Malaysia Directory* is the only engine that allows Boolean and keyword or phrase searching. Other than that, *Malaysia Directory* also focuses on other portal features that provide services such as “Work Group Mail Server”, “Web and Email hosting services”, “Commercials” and “Hobby Communities”, which are not commonly offered by the other three search engines. *Malaysia Directory* and *Cari* top the list in terms of additional portal features.

Hananzita, H. & Kiran K.

Google's key feature is simplicity, in which it offers clean design that gives clear direction on how user can proceed in using its search engine. The only local engine that emulates this type of feature is *SajaSearch*. It offers simple and direct feature and interface. There are no advertisements or banners or any other portal features that can sometimes be annoying. Similarly, *Malaysia Central* also offers simplicity with minimal graphic environment feature. It seems that these two engines i.e. *SajaSearch* and *Malaysia Central* are more concerned about search functionalities rather than colorful and impressive design as being offered by *Cari* or *Malaysia Directory*.

Retrieval Performance

A total of twenty-five (25) predetermined queries were submitted to each engine. These queries were selected based on popular terms/phrases used by *Yahoo!Malaysia* and *Catcha Malaysia*. Refer to Appendix for list of queries. Queries were posed in two different languages i.e. English and Bahasa Melayu in order to assess the search engine capability to produce results in those languages. Query results returned by Google were not included in the comparison because of the huge difference of number. Figure 2 shows the total number of documents returned for the given queries submitted to the engines. It is to be noted that the total number of result here contained redundant documents retrieved by the four search engines. Due to the vast number of difference of total hits, the benchmark engine, Google, was not included in the comparison. Of the four search engines, it is clear that *Malaysia Directory* rarely produced many results, mostly with zero retrieval, and often when it did find matches, the number of broken links was very high. Whereas, *SajaSearch* almost always returned results for any queries submitted to the engine. However, *Cari* produced the highest total hits for the given queries. The figure is followed closely by *SajaSearch* with only a difference total number of 3 of total hits.

In general, out of the four engines, *SajaSearch* heads the list in returning the most related documents out of all total documents retrieved. This is closely followed by *Cari*. Even though *Cari* retrieved the most total hits (389 documents), it shows that only 81 documents were related. *SajaSearch* managed to retrieve 85 related documents out of 386 documents. However, even though it is the highest figure, still it is considered very low in percentage, which is about only 22% of the whole total documents retrieved by the four engines. Interestingly, if we compare which engine returned the highest percentage of related documents out of the total documents that were retrieved solely by the particular engine, *Malaysia Central*

scored highest. Out of 143 documents, *Malaysia Central* was able to retrieve 62 related documents, which is about 43% of its total documents. Obviously, *Malaysia Central* ranked highest in term of relevancy from this perspective (Figure 3).

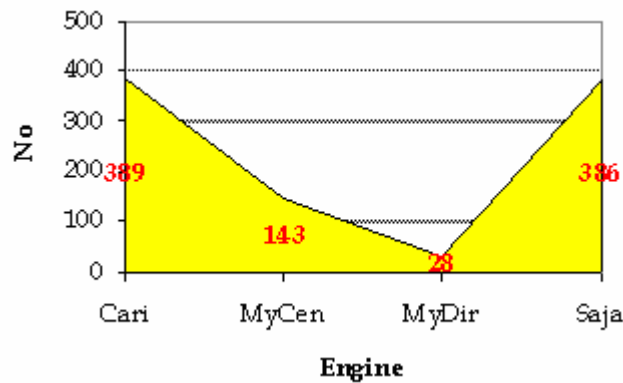


Figure 2: Total Documents Retrieved

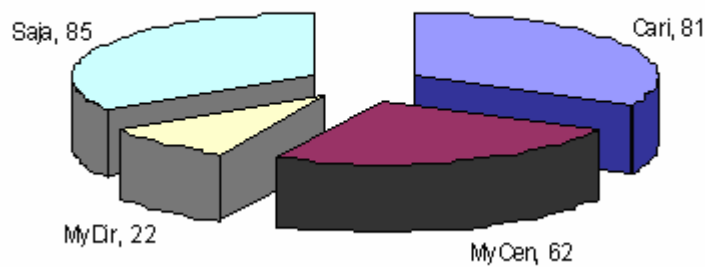


Figure 3: Relevant Documents Retrieved

Figure 4 shows the total hits for each query (25 in all) submitted to every search engine. Query #20: “SPM”, returned the most total hits compared to other queries. This is followed by Query #12: “Siti Nurhaliza” with total hits of 146. The reason underlying the most and least total hits results are simply because of the currency state of the queries. Query #20: “SPM” produces the most total hits because it is an old, on-going term that is well-known and commonly used in Malaysia. Whereas,

there are two queries which are Query #5: “tsunami” and Query #25: “perpustakaan digital”, returned least total hits result since these terms are genuinely ‘new’ terms for Malaysians. Clearly enough, there are not many documents containing these terms being indexed by the four Malaysian Web search engines.

SajaSearch outperforms other search engines with no zero retrieval, the only engine that can closely outdo the benchmark engine, Google. Even with good search features, *Malaysia Directory* performs badly with highest zero retrieval, which means this engine only able to return results for seven (7) out of twenty five (25) queries submitted to the engine (Figure 5). Although *SajaSearch* outperforms in terms of most related documents, and with no zero retrieval, this engine scored second to last in terms of mean precision. *Malaysia Central* surprisingly scored highest for mean precision, with only 0.01 differences with *Cari*. Figure 6 illustrates the difference of mean precision among the selected search engines with Google.

The best search engine based on precision score is *Malaysia Central* and the worst is *Malaysia Directory*. Precision is congruently related with relevance. The higher number of relevant documents for the total number of documents retrieved is, than the higher value of precision will be. With this reason, it clearly explains on why *Malaysia Central* leads in the precision score if compared to the other engines. The same goes for why *Malaysia Directory* scored the least precision.

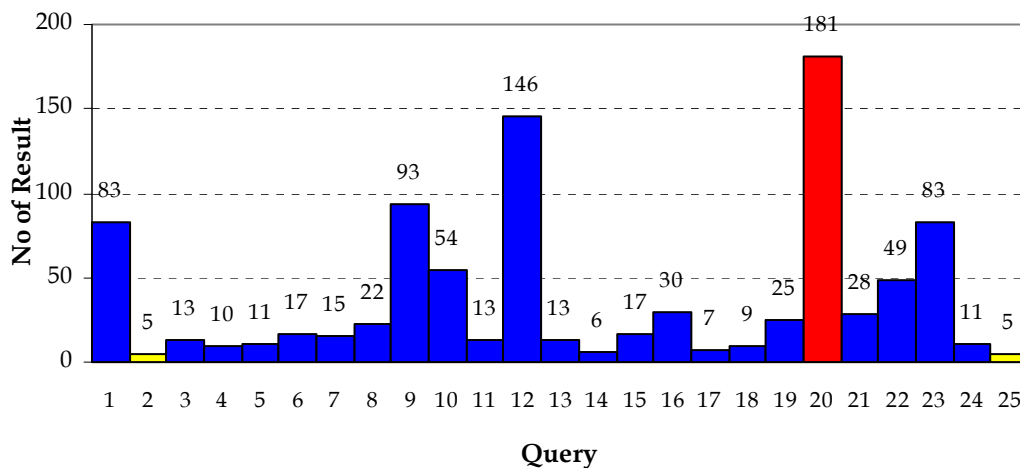


Figure 4: Query Result

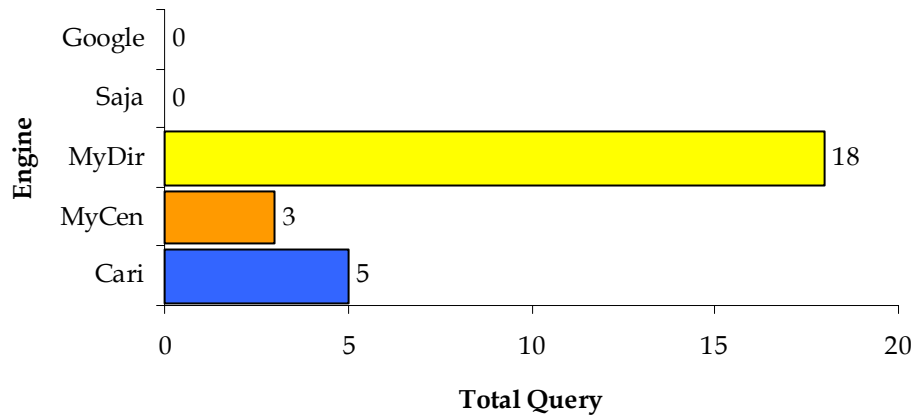


Figure 5: Query with Zero Retrieval

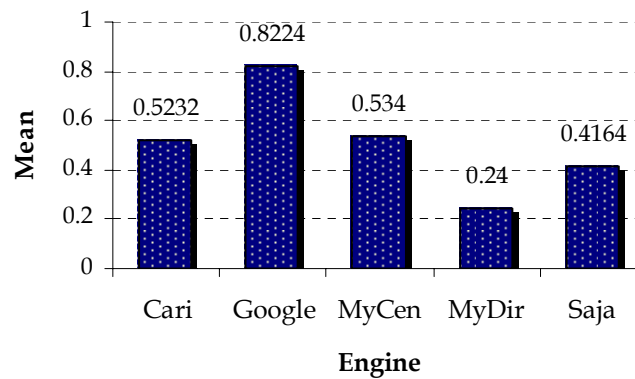


Figure 6: Mean Precision between Selected Search Engines and Google

It is quite difficult to explain reason why *Malaysia Central* is able to find more relevant documents compared to others. The reason underlying the performance is actually difficult to tell without fully knowing the design of each search engine and how the indexes are created. Normally the designs are considered as trade secrets and will never be published. The conclusion for the best search engine according to its category is presented in Table 3 based on the overall comparison of the selected search engines (Figure 7).

Hananzita, H. & Kiran K.

Apparently, no engine is the best engine in all of the categories. The engine with the best feature, *Malaysia Directory*, shockingly scored lowest in other categories. This shows how a good search features and presentation does not necessarily means the engine is good in terms of performance. Likewise, the engines that can produce the most total hits and related documents do not score well in the mean precision. *SajaSearch* scored highest for the engine that retrieved most related documents from the total whole retrieved documents. However, as previously being discussed, *Malaysia Central* is better at producing related documents for documents solely retrieved by its engine. Consequently, this also clarifies the reason for *Malaysia Central* is the best engine with highest mean precision score.

Table 3: Best Search Engine According to Category

| Category | Search Engine |
|------------------------|------------------------------------|
| Best Feature | <i>Malaysia Directory</i> |
| Best Total Hits | <i>Cari</i> |
| Best Related Documents | <i>SajaSearch/Malaysia Central</i> |
| Best Mean Precision | <i>Malaysia Central</i> |

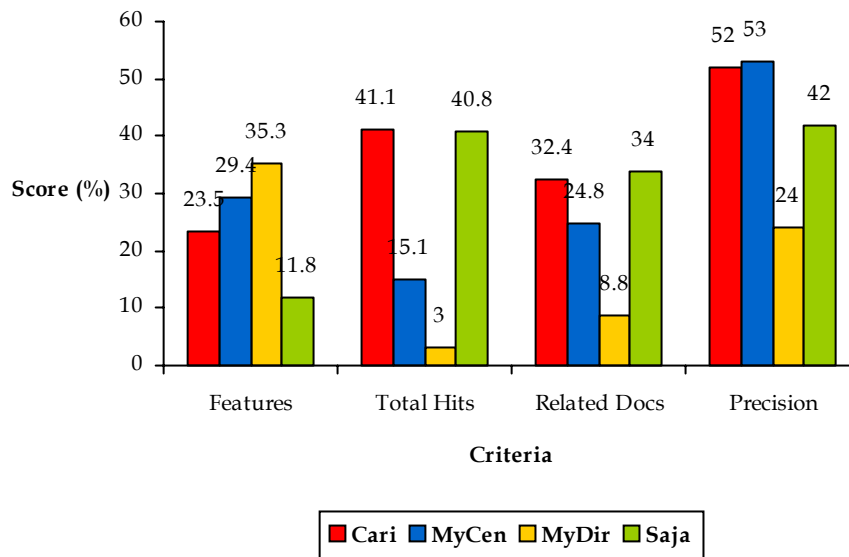


Figure 7: Overall Comparison of Selected Search Engines

Comparison With Benchmark Engine

Each engine that scored highest in its category is compared with the benchmark engine, Google and the result is displayed in Table 4. There are still huge gaps in scores if local search engines are compared to the benchmark engine, Google. *Cari*, with the highest score for total hits returned 389 documents and this differs greatly with the total hits returned by Google. In total, Google actually returned more than half a million of documents for the queries submitted to the engine, but for this research purposes only the first twenty results are focused, thus the total hits is 500 instead of 666, 168. The total number of documents of all the four Malaysian Web search engines combined was unable to outperform the total numbers of document retrieved by Google’s actual figure. The number of related documents shows a huge number of discrepancy in which *SajaSearch* only retrieved 85 related documents whereas Google retrieved almost five times more with a total of 410. The same goes for mean precision in which Google’s performance reveals enormous difference compared to even the most precise local search engine studied.

Table 4: Comparison Malaysian Engines with Google

| Evaluation Criteria | Engine with Highest Score | Google |
|----------------------------|----------------------------------|---------------|
| Features | <i>Malaysia Directory</i> = 6 | 9 |
| Total Hits | <i>Cari</i> = 389 | 500(666,168) |
| Related Documents | <i>SajaSearch</i> = 85 | 410 |
| Mean Precision | <i>Malaysia Central</i> = 0.53 | 0.82 |

Strength and Weakness of Each Engine

Each of the engines portrays different features and capabilities. Each one of them possesses their own strengths and weaknesses. Table 5 summarizes their strengths and weaknesses. Overall, the major limitation of the local Web search engines is in term of their searching features. None of the engines fully support Boolean searching, or nesting as well as truncation. Some do not provide any assistance for users on how to use their engines. Even the most popular engine in Malaysia, *Cari*, does not offer help or any additional search features for users.

Table 5: Strength and Weakness of Each Search Engine

| Search Engine | Strength | Weakness |
|---------------------------|--|---|
| <i>Cari</i> | <ul style="list-style-type: none"> <input type="checkbox"/> Most popular Website compared to other local engines – ranked no. 7 in Malaysia Top 5 Website <input type="checkbox"/> Produced the highest total hits <input type="checkbox"/> User-friendly interface | <ul style="list-style-type: none"> <input type="checkbox"/> Produced most duplicate documents <input type="checkbox"/> Limited search features: no nesting, no truncation, does not support full Boolean <input type="checkbox"/> No user aids <input type="checkbox"/> Page full of advertisement, banners <input type="checkbox"/> Only retrieve web files (.html) |
| <i>Malaysia Central</i> | <ul style="list-style-type: none"> <input type="checkbox"/> Highest score for mean precision <input type="checkbox"/> Advertisement-free, no banners, no pop-ups and minimal graphics environment <input type="checkbox"/> Does not imitate Yahoo!’s directory layout <input type="checkbox"/> Search results carry physical address, phone and fax numbers <input type="checkbox"/> Produce least duplicate documents <input type="checkbox"/> User-friendly interface | <ul style="list-style-type: none"> <input type="checkbox"/> Limited search features: no nesting, no truncation, does not support full Boolean <input type="checkbox"/> Only retrieve web files (.html) |
| <i>Malaysia Directory</i> | <ul style="list-style-type: none"> <input type="checkbox"/> Provide search option feature that support Boolean and keyword or phrase search <input type="checkbox"/> User-friendly interface | <ul style="list-style-type: none"> <input type="checkbox"/> Produced most zero retrieval <input type="checkbox"/> Produced least related documents <input type="checkbox"/> Page full of advertisements and banners <input type="checkbox"/> Only retrieve web files (.html) |
| <i>SajaSearch</i> | <ul style="list-style-type: none"> <input type="checkbox"/> Produced most related and updated documents <input type="checkbox"/> Shows searching length time <input type="checkbox"/> Interface similar to Google <input type="checkbox"/> Advertisement-free, no banners, no pop-ups and minimal graphics environment <input type="checkbox"/> Support other file types other than html – PDF file <input type="checkbox"/> No zero retrieval <input type="checkbox"/> User-friendly interface | <ul style="list-style-type: none"> <input type="checkbox"/> Limited search features: no nesting, no truncation, does not support full Boolean <input type="checkbox"/> No user aids |

CONCLUSION

In a nutshell, the performance and features of local Web search engines greatly differ in various aspects from the well-established commercial search engine, Google. Based on the findings discussed, none of the local search engines fulfill all the evaluation criteria that have been drawn in this study. However, relevance and

Malaysian Web Search Engines

precision are the key factor in attracting users to keep coming back to use the search engine as the starting point of a search for information. In the context of library and information science field, providing relevant materials in response to a user's query, is the main goal. As a consequence, a search engine that returns a high number of relevant documents is better than a search engine that provides good features but returns zero retrievals. A capability of a search engine is meaningless if it returns zero retrieval, no matter how impressive and advanced the search features are. Simplicity as shown by *SajaSearch* is more essential in creating user's interest to use the search engine. *SajaSearch* has proven that it need not be impressive in terms of appearance in order to provide a good list of related documents.

Being popular also does not necessarily means that the search engine is good. *Cari*, being the most popular among the search engines being studied, does not reveal it as being the best search engine among those four engines. It only wins in terms of returning the most total hits. Even though the total hits returned reveals quite an impressive figure, the number of related documents was not great enough to be considered as the best search engine. *SajaSearch* outperforms *Cari* in terms of most related documents, while in terms of precision, *Malaysia Central* outperforms *Cari*.

There is still in need for a truly 'localized' search engine that completely indexes and searches only Malaysian Web sites. The existing local search engines still require improvements and upgrading in order to overcome the limitations and shortcomings and comparable to the outstanding benchmark engine like Google. Currently, the Malaysian Web search engines search for documents or links that contain information rather for the information itself. It would be a user's wish list to have a search engine that can search for information not documents. Further studies and improvement to build fast, accurate and reliable local search engines that can overcome the lack of relevant search results are crucial indeed. This is actually also applied to other than Malaysian Web search engines.

REFERENCES

- Abdul Samad, R. 2001. The double edged sword: A brief comparison of Information Technology and Internet development in Malaysia and some neighbouring countries. *IFLA Journal* 27 (2001): 315-318. Available at: <http://wotan.lib.edu/doi/data/Articles/julksrngay:2001:v:1:5-6?p.314-319.html>
- Botluk, D. 2000. *Update to search engines compared*. Washington: Catholic Univ. of America.

Hananzita, H. & Kiran K.

- Chowdhury, A. and I. Soboroff. 2002. *Automatic evaluation of World Wide Web search services*. Tampere, Finland: SIGIR '02.
- Chu, H. and Marilyn Rosenthal. 1996. Search engines for the World Wide Web: A comparative and evaluation methodology. *ASIS '96 Annual Conference Proceedings* 1996. Available at: <http://www.asis.org/annual-96/ElectronicProceedings/chu.html>
- Davis, E. T. 1996. *A comparison of seven search engines*. Available at: <http://www.iwaynet.net/~lsci/search/paper-only.html>
- Google History. *Google corporate information*. 2004. Available at: <http://www.google.com/corporate/history.html>
- Gwizdka, J., and M. Chignell. 1999. *Towards information retrieval measures for evaluation of Web search engines*. Toronto, Canada: University of Toronto.
- Hashim, R., and A. Yusof. 1998. Background of Internet development in Malaysia. *Diffusion of Internet: Preliminary findings in Malaysia*. Available at: http://www.interasia.org/malaysia/preliminary_background.html
- Hawking, D., N. Craswell, P. Thistlewaite, and D. Harman. 1999. Results and challenges in web search evaluation. *Proceedings of WWW8*. 1999. Available at: <http://www8.org/w--papers/2csearch-discover/results/results.html>
- Lawrence, S. and Giles, C. L. 1998. Searching the World Wide Web, *Science*, Vol. 280, no. 5360: pp. 98-100.
- Leighton, H. V., and J. Srivastava. 1997. Precision among World Wide Web search services (search engines): Alta Vista, Excite, Hotbot, Infoseek, Lycos. Available at: <http://www.winona.msus.edu/library/webind2/webind2.htm>
- Leighton, H.V. 1996. *Performance of four World Wide Web (WWW) Index Services: Infoseek, Lycos, Webcrawler and WWWorm*. Available at: <http://www.winona.edu/library/webind.htm>
- Loose, R.M. and L. A. H. Paris. 1999. Measuring search engine quality and query difficulty: Ranking with target and freestyle. *Journal of the American Society for Information Science*, Vol. 50: 882-889. Available at: <http://portal.acm.org/citation.cfm?id=318976.318982>
- Malaysia Internet Usage and Marketing Report. *Internet World Stat*. 2004. Available at: <http://www.internetworldstats.com/asia/my.htm>
- Malaysia Internet Usage and Marketing Report. *Internet World Stat*. 2005. Available at: <http://www.internetworldstats.com/asia/my.htm>
- Malaysia Top 50 Most Popular Websites. *Criteria to be included in the Malaysia Top 50*. 2005. Available at: <http://www.Cari.com.my/top50/>
- Ohtsuka, T., K. Eguchi, and H. Yamana. 2004. *an evaluation method of Web search engines based on users' sense*. Tokyo, Japan: National Institute of Informatics.
- Rijsbergen, C.J. 1979. *Information retrieval*. 2nd ed. London: Butterworths.

Malaysian Web Search Engines

- Shang, Yi and L. Li. 2002. *Precision evaluation of search engines*. Netherlands: Kluwer Academic. (PDF version of document downloaded 6 December 2004).
- Sugiyama, K., K. Hatano, and S. Uemura. 2004. *Adaptive Web search based on user's implicit preference*. Nara, Japan: Nara Institute of Science and Technology. (PDF version of document downloaded 12 January 2005).
- Sutachun, T. 2000. *A comparative study of Internet search engines*. Available at: <http://websites.kku.ac.th/abstract/thesis/mart/lis/2543/lis430009t.html>
- Tang, M.C. and Y.Sun. 2000. *Evaluation of Web search engines using user-effort measures*. Available at: <http://libres.curtin.edu.au/libres13n2/tang.htm>
- Top 20 Countries with The Highest Number of Internet Users. *Internet World Stat: Usage and population statistics*. 2004. Available at: <http://www.internetworldstats.com/top20.htm>
- Wang, Z. 2001. *Information systems and evaluations. Precision and recall graphs*. Available at: <http://www2.sis.pitt.edu/~ir/Projects/Sum01/FinalProjectsSum/ZhangWang/ZhangWang-finalprojecthandin/List.htm>
- Wishard, L. 1998. *Precision among Internet search engines: An earth science case study*. Available at: <http://www.library.ucsb/ist/98-spring/article5.html>

APPENDIX

List of Queries Submitted to Selected Search Engines and Google

| Sample Queries | Google | Other Engines |
|----------------|---------------------------------|-------------------------------|
| Query #1 | “Chinese New Year” | Chinese New Year |
| Query #2 | denggi | denggi |
| Query #3 | “virus denggi” | virus denggi |
| Query #4 | “wabak demam denggi” | wabak demam denggi |
| Query #5 | tsunami | tsunami |
| Query #6 | “trauma tsunami” | trauma tsunami |
| Query #7 | “pembebasan Anwar Ibrahim” | pembebasan Anwar Ibrahim |
| Query #8 | “pendatang asing” | pendatang asing |
| Query #9 | “Le tour de Langkawi” | Le tour de Langkawi |
| Query #10 | resipi | resipi |
| Query #11 | “violence against women” | violence women |
| Query #12 | “Siti Nurhaliza” | Siti Nurhaliza |
| Query #13 | MyKad | MyKad |
| Query #14 | “Naza Kia” | Naza Kia |
| Query #15 | ”Naza Citra” | Naza Citra |
| Query #16 | “smart school” | smart school |
| Query #17 | “selesema burung” | selesema burung |
| Query #18 | “Suruhanjaya Perkhidmatan Awam” | Suruhanjaya Perkhidmatan Awam |
| Query #19 | “khidmat negara” | khidmat negara |
| Query #20 | SPM | SPM |
| Query #21 | “sejarah Melayu” | sejarah Melayu |
| Query #22 | “kota gelanggi” | kota gelanggi |
| Query #23 | recipes | recipes |
| Query #24 | cybercrime | cybercrime |
| Query #25 | “perpustakaan digital” | perpustakaan digital |