

## A quality analysis of keyword searching in different search engines projects

C. Wu<sup>a</sup>, K. Jenab<sup>b\*</sup>, S. Khoury<sup>c</sup> and S. Moslehpour<sup>d</sup>

<sup>a</sup>Graduate Student, Dept. of ETM, Morehead State University, KY, USA

<sup>b</sup>Faculty of Dept. of ETM, Morehead State University, KY, USA

<sup>c</sup>Graduate Program Director, Coordinator, Computer Information Systems, Division of Business, Spring Hill College, Mobile, AL, USA

<sup>d</sup>Professor of Electrical and Computer Engineering, University of Hartford, Hartford, CT, USA

### CHRONICLE

### ABSTRACT

#### Article history:

Received: September 30, 2017

Received in revised format: October 10, 2017

Accepted: December 5, 2017

Available online:

January 2, 2018

#### Keywords:

Quality

Keyword Searching

Search Engine

A search engine is an essential tool in our daily life. With the development of society and network technology, the users' requirement of Internet information is increasing. For most search methods, keyword searching is in a crucial position. However, what about the quality of keyword search in different search engines? This paper evaluates the quality of keyword searching among different search engines project.

© 2018 by the authors; licensee Growing Science, Canada.

## 1. Introduction

Internet information is becoming more and more essential to people. The effective search tool is receiving more attention by researchers than ever before. With the development of society and network technology, the users' requirement of Internet information is increasing. From a seemingly unlimited knowledge reservoir, the search engines (SEs) can help people meet the required information by inputting some keywords. Different users have different needs. They can choose diverse search tools to reach their requirements. The main difference between basic SEs and special SEs is the various additional features special SEs provide in addition to those offered by basic SEs. Although most people will choose the basic SEs first, the special SEs can meet the special needs of the Internet information. The basic SEs in combination with some special search features can meet the users' needs, but it is not an effective way to reach the final requirement. The better way to get diverse needed information is to use the special SEs. In addition, the very important thing is that there is a difference in quality between them. Therefore, users will benefit from the identification of those features that provide maximum quality. In this study, the researchers distinguish between two different types of SEs, which are vertical SEs and comprehensive SEs. Then, the researchers use four SEs and divide them into two categories,

\* Corresponding author. +1-606-783-9339

E-mail address: [k.jenab@moreheadstate.edu](mailto:k.jenab@moreheadstate.edu) (K. Jenab)

which are Google and Baidu as comprehensive SEs and Amazon and JD.COM as vertical SEs, to compare the search features via searching keywords and to discuss the quality of each SE. This study reports the SEs' quality and their fitness for use by users.

## 2. Literature Review

Search Engines (SEs) are tools that can help users find related information via input of keywords or phrases. They are also computer programs that meet users' diverse information needs. SEs compare the search words with a webpage content index file. The results are then returned to the user's screen (Weideman, 2004). Users usually enter the keywords into the search box to retrieve information from the Internet. The overall popularity of a website is determined by the "link popularity" and "click popularity", two factors that influence the ranking of the website. The SE selects an array of webpages to determine which pages are most relevant. These webpages contain some of the queried items. Then, the SE will calculate a score for each webpage and produces a list of webpages sorted by the SEs scoring system (Egele et al., 2009). People always use the SEs to search information on the Internet (Blumauer & Hochmeister, 2003). These SE indexes are usually made by human editing or updated by computer programs called spiders (Weideman, 2004). SEs use a variety of complex algorithms to check the value of web content for the user. Furthermore, they use "spiders" to find keywords and to locate readable content within webpages (Ramos & Cota, 2004). Of these SEs, the latest figures show that Google dominates the market at 66.4% of the market share. (Sterling, 2012).

According to the user's search behavior, we should determine the best measure in terms of the number of words: it should be enough to get a large number of keywords, but not too much (Visser, & Weideman, 2011). Previous research showed that if there are keywords in the title and in the body of the webpage, the SE would get a better result (Zhang & Dimitroff, 2005). Keywords search supported by structured data is beneficial, since it provides richer semantics than text documents. This provides us with better opportunities to generate high-quality results (Termehchy & Winslett, 2009). As evident in the existing literature, by comparing the features of SEs through different points of view and diverse ways, the conclusion shows mixed or contradictory results (Robinson & Wusteman, 2007; Hochstotter, & Koch, 2009; Uyar, 2009).

## 3. Methodology

The purpose of this study is to investigate the quality of SEs responses to users' keyword searching and to record users' opinions in relation to different kinds of SEs and the retrieved results. There were two main goals of this research to:

- Evaluate how well vertical and comprehensive SEs respond to keyword searching; and
- Assess whether the vertical and comprehensive SEs are more effective in satisfying user information needs.

Therefore, the research question is, "Do vertical and comprehensive SEs perform good quality in keyword searching and are they successful in satisfying user information needs?" This research uses a comparison methodology. Four SEs are selected, because of their popularity amongst users and because they represented two different types of SEs, which are vertical and comprehensive SEs. The four SEs are: Google.com, Baidu.com, Amazon.com, and JD.COM. The subjects are based on real user information needs. However, each keyword search is used independently from the entirety of the information need. The research is constructed and divided into four quality sections: Search completion time, Number of webpages shown in a search task, Precision, and Relative Recall. For each search task, ten keywords are submitted to the four SEs using the above-mentioned four quality features. By contrasting the results of the data, the researchers are able to find the answer to the research question.

## 4. Definitions

### 4.1 Search Engine (SE)

A SE is a system that uses a specific computer program to collect information from the Internet. A SE is not only a necessary function for users, but also an effective tool for the behavior of a web user. The efficient SE allows the user to find target information accurately and fast (Antriksha, & Ugrasen, 2011). The search results are usually shown in a series of results, usually called SE result pages. The types of shown information are always different, which include webpages, images, and other types of files. Some SEs can also get important data in the database or open directories. SEs can also run an algorithm on a crawler to maintain the information in real-time from different web directories. The information that a search processes should have high precision and meet the requirement of the user. After generating the search results, the ideal SE should have both a simple query and advanced search functions at the same time (Meng & Songyun, 2011). Different types of SEs that are readily available can address differences in information collecting methods and services.

### 4.2 Keyword

A keyword, refers to specific words that express the webpage features. Keywords are used as shortcuts which sum up an entire page. As the component of the metadata of the webpage, keywords help SES match an appropriate search query. Keywords become important in SES because they make connection between the content of the webpage and user's inquiry.

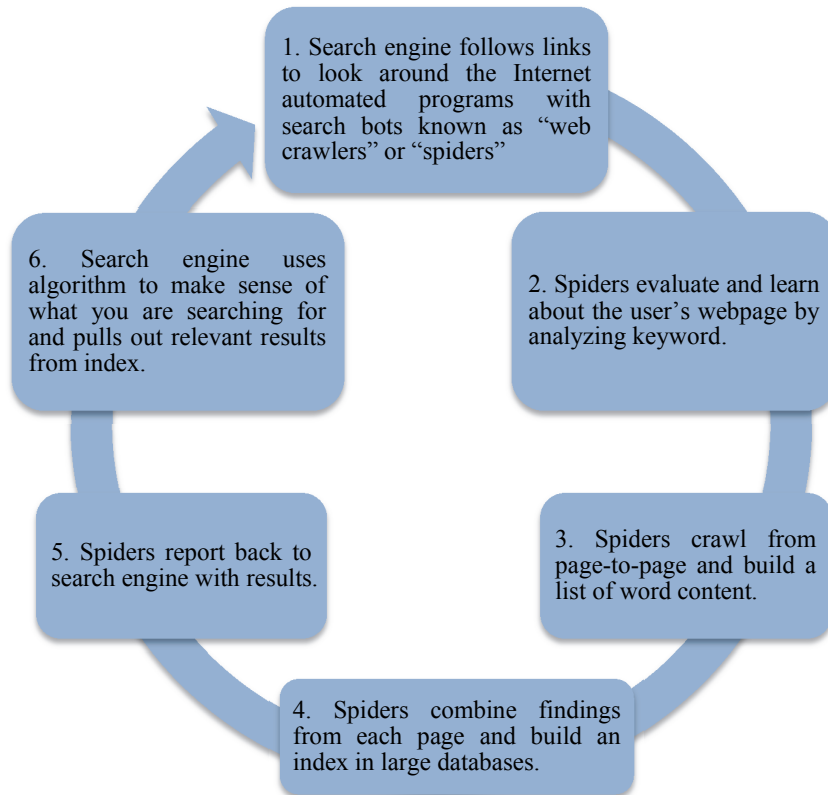
### 4.3 How does the SE work by keyword?

The SE deals with tens of thousands of information searches. The process follows the pre-determined rules of the SEs' operating principles. SEs will request information according to the three following steps (Meng & Songyun, 2011):

- 1) Crawl Page: Each individual SE has its own web capture process. It, along with the hyperlinks of the web, continuously capture the pages. The capture page is called a webpage snapshot. Due to the application of hyperlinked Internet pages, theoretically, starting from a range of webpages, we can collect the vast majority of pages that are related to our keyword.
- 2) Processing Page: After catching webpages, SEs still need to do lots of pre-treatment projects to provide retrieval service, among them, the most important part is extracting keywords and establishing index files. Others include removing duplicate webpages, participles, judging the types, analyzing hyperlinks and counting pages important degree/abundance, etc.
- 3) Providing Search Services: User inputs the keywords then the SE finds the matching pages from the indexed database; except for page, title and URL, it still provides an abstract from webpages and other information to make user's estimate expediently. The work process of SE is shown in Fig. 1.

These huge storage devices enable thousands of machines to process much information quickly. When a person searches on any major engine, they request the result immediately; even a one- or two-second delay will cause users' dissatisfaction, so the SE must provide the answer as quickly as possible.

The most useful feature of a SE is the relevance of the returned result set. Although there are millions of webpages, which include a specific word or phrase, some of them may be more relevant, popular, or authoritative than others. Most SEs use methods to sort and provide the best results.



**Fig. 1.** How does Search Engine work?

## 5. Search Engines (SEs)

### 5.1 Types of SEs:

A SE is one of the most important tools of information service on the Internet. Although it has seen much improvement in recent years, its service functions received the most attention. In this paper, the SEs are classified into two types, one is the vertical SE and the other one is the comprehensive SE. Comprehensive SE is defined relative to the vertical SE, and it is our traditional SE. The search resources are exhaustive, and users can input a keyword to recall almost any type and any subject resources. It is most useful when looking for specific sites or very unique subjects and can satisfy users' requirements for massive information. However, there are some disadvantages. First, it is very difficult to get higher accuracy and relevancy of search quality with thousands of irrelevant results. Second, there are many dead links and low correlation links. Lastly, for the special customer's requirement, there are no clear directions to get more detailed and centralized information. The different comprehensive SEs are shown in Table 1.

A vertical SE collects web information from multiple and different resources in a specific domain, and reorganizes them as structured data, so it can provide more professional and individualized information services for special customers and satisfy their requests for detailed information in their domain (Wu et al., 2010). The application of vertical SEs is broad, such as job search, tourism search, medical search, book search, shopping search, and so on. It can be further refined into various kinds of vertical SEs in every walk of life. The different vertical SEs are shown in the Table 1.

**Table 1****Different SEs of Comprehensive and Vertical SEs**

	Comprehensive Search Engines	Vertical Search Engines
1	Google	Amazon
2	Bing	Alibaba
3	Baidu	Taobao
4	Yahoo!	JD.COM
5	Ask	Youtube
6	Aol search	Bestbuy
7	DuckDuckGo	Ebay
8	Dogpile Search	Facebook
9	Wolfram Alpha	Kayak
10	Webopedia Search	Yelp

*5.2 Features of SEs*

For *comprehensive SEs*,

1. It provides a search entrance to search the related questions of users in different webpages. Then users find out the related information and they must determine the relevance of information. The keyword must be complicated and users must search the clear requirement of the information.
2. The search results are webpage links, and the principle of search is the description of webpages and relevance of keywords.
3. It depends on the search system algorithm and the results are an automatic page arrangement. Users cannot choose the arrangement and only accept the order of the SE.
4. In the search results, they are described including three parts: title, description, and URL link. These descriptions are more about the introduction of the overall content of the webpage on the current URL link, rather than the specific information the user searches.
5. The results are often a huge number of webpages. So, the recall ratio is high. But, the SE is searching from the extensive Internet searching and the user cannot find the results very accurate. Therefore, the precision ratio is relatively low.

For *vertical SEs*,

1. Users have a clear demand for information, and the need of information can be defined in a specific range. The information product is in a specific form and organization. The users do not have to carry on the analysis and judgment for information. The users just need to search the easy keyword and the results are precise.
2. Its search results are structured data, it almost has no need for users to specifically open webpages, and they can determine whether the results are their own results.
3. The arrangement can be set by the users and they can independently choose the arrangement according to the relevance ranking, the price, the scope of the price, and other ways. It is helpful for the users to find the information that users need.
4. It has strong pertinence in the search results and describes the specific information that users look for from multiple aspects. Users do not need to click the link directly to determine which search results are the most needed information.
5. The results are limited. So, the recall ratio is low. But, the SE is searching from a particular website and the user can find the accurate results. Therefore, the precision ratio is very high.

The comparison of features between comprehensive and vertical SEs is shown as Table 2.

**Table 2**

Comparison of Features between Comprehensive and Vertical SEs

	Comprehensive search engine	Vertical search engine
Form of search results	Simple description and link of webpage	Structured data
Arrangement of search results	Systematic algorithm	Setting by users
Description of search results	Huge amount	Limited
Recall ratio of search results	Relatively low	High
Precision ratio of search results	Title, description, URL link	All the information related to the

## 6. Introduction of Different SEs

### 6.1 Comprehensive SEs: Using Google and Baidu as example

#### *Google Search*

Google Search, commonly referred to as Google Web Search or simply Google, is a web SE developed by Google. It is the most-used SE on the World Wide Web, handling more than three billion searches each day (Burns, 2008). As of February 2016, it is the most used SE in the US with 64.0% market share (Burns, 2008). The order of search on Google's search-results pages is based, in part, on a priority rank called a "PageRank". Google Search provides many different options for customized search, using Boolean operators. Google uses an algorithm, but its algorithm is based on answering user search queries. To this end, Google relies on user engagement and external trust factors for judging the relevancy of a search result. Google calculates SE Optimization (SEO) using a range of on-page factors including session duration, bounce rate, click-through-rate, etc. as well as off-page factors including social mentions, quality backlinks, and domain authority (Burns, 2008).

#### *Baidu Search*

Baidu is a dominant Chinese Internet SE company. It offers many of the same products and services as Google, but is primarily focused on China, where it controls most of the search market. Baidu censors search results and other content in accordance with Chinese regulations. Baidu presents several keyword-based discussion forum (Jiang, 2014). Baidu has the 2nd largest SE in the world, and held a 76.05% market share in China's SE market, the largest in the world, as of April 2017. As of 2017, Baidu Search released Spider 3.0, which is capable of indexing over trillions of webpages. Baidu maintains by far the biggest share of the SE market in China. Besides being an early mover, one of the main reasons for Baidu dominating the market is their ability to parse and interpret Chinese text more effectively than other SEs, leading to higher-quality results. The SE gives much higher priority to Chinese language sites, and indexes far fewer non-Chinese language sites (Jiang, 2012).

### 6.2 Vertical SEs: Using Amazon and JD.COM as example

#### *Amazon*

Amazon is an American electronic commerce and cloud computing company based in Seattle, Washington that was founded by Jeff Bezos on July 5, 1994. They are the second largest Internet retailer, coming in just under alibaba.com (Jopson, 2011). Amazon uses the A9 search algorithm to locate relevant products for its users. A9 has development efforts in areas of product search, cloud search, advertising technology, and community question answering. It does this by considering "human judgments, programmatic analysis, key business metrics, and performance metrics." The focus of Amazon's SE is finding and displaying products that have a high conversion (sales) rate. Amazon judges search relevancy by on-page factors like product sales and availability, customer reviews, price, image size/quality, and related products. Notice that all of these factors are included on the product page itself, not through backlinks or social media platforms (Jopson, 2011).

Amazon's product listings rely on individual keywords, not key phrases. Words listed in the product title, brand, etc. are automatically counted as keywords and do not need to be repeated in the product description or in the search term fields. Amazon relies on results and conversions when ranking products. The more customer reviews and sales our products generate, the more prominently our products will get ranked by Amazon, initiating a self-perpetuating cycle of more conversions=better rank=more conversions (Jopson, 2011).

## *JD.COM*

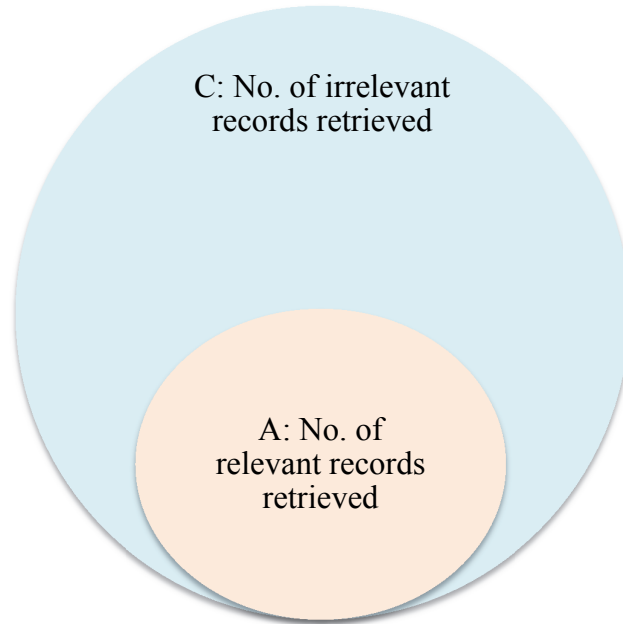
JD.COM is a url for Jingdong located in Beijing that is formerly called 360buy. Considering transaction volume and revenue, Jingdong is one of the two largest Business to Commerce (B2C) online retailers in China. Also, it is a member of the Fortune Global 500, and is a major competitor to Alibaba-run Taobao. Currently, it has 258.3 million monthly active users (*JD.com*, 2017). JD.COM is the world's leading company in high tech and AI delivery through drones, autonomous technology and robots, and possesses the largest drone delivery system, infrastructure, and capability in the world. It has recently started testing robotic delivery services and building drone delivery airports, as well as operating driverless delivery by unveiling its first autonomous truck (*JD.com*, 2017). JD.COM has formed a strategic partnership with Chinese SE Sogou, to leverage big data to improve targeting. The move comes months after the e-commerce giant inked a similar deal with search powerhouse Baidu in a bid to help brands target consumers more effectively. The deal will give Sogou users direct access to JD.COM's shopping platform via Sogou's search, news aggregation, and yellow pages mobile apps. Sogou, which is a subsidiary of one of China's leading online media, video, search, and gaming business group Sohu, is the latest technology company to partner with JD.COM, which is on a mission to boost its brand and services as it competes with Alibaba. Baidu, China's largest SE, has struck a deal to funnel users looking for products to online retailer JD.COM (*JD.com inks partnership deals with Chinese search engine Sogou*, 2017)

## **7. Quality Analysis of SE**

### *7.1 Quality Criteria*

High quality sites should provide positive experiences for the visitor. In this paper, the research of quality is divided into four quality sections: Search completion time, Number of webpages shown in a search task, Precision, and Relative Recall.

1. Search completion time: It is a calculated amount of time required for any particular task to be completed. This is a typical metric in usability evaluation. During this research, users were told to read the task and then to click a "start searching" button, which would begin the search session by opening the appropriate search algorithm (Ya & David, 2009). When the results are shown in the searching webpage, the search task is finished.
2. Number of webpages shown in a search task: is the number of unique SEs or databases used by a participant in a task. When the users search some keyword, the number of webpages is shown in the result of the searching webpage.
3. Precision: It is usually expressed as a percentage that is computed by Equation (1) (Tauqeer, 2012). The composition of a search record is shown in Fig. 2.



**Fig. 2.** The composition of Search Record

$$\text{Precision} = \frac{A}{A + C} \times 100\% \quad (1)$$

4. Relative Recall: is usually expressed as a percentage. We can calculate it via dividing the total number of all relevant records in the database by the number of relevant records retrieved (Tauqeer, 2012).

$$\text{Relative Recall} = \frac{\text{Number of sites retrieved by search engine}}{\text{Total number of sites retrieved by all search engine}} \quad (2)$$

## 7.2 Comparisons of Different SEs in Quality Criteria

In this section, the researchers randomly choose ten different keywords. Five of the keywords are selected from the 100 most popular Google keywords (*“the 100 most popular Google keywords”*, 2017) and the other five keywords are selected from Top Baidu Searches 2016 (*“Top Baidu Searches 2016”*, 2016). The five Google keywords are “weather”, “translate”, “maps”, “news”, and “calculator”. The five Baidu keywords are “QQ”, “G20”, “Alipay”, “Wechat” and “IQiYi”. In this paper, the researchers chose a tool from Chrome called tools for web developers to evaluate the search completion time. Each keywords search is depended on this tool.

### 7.2.1 Comprehensive SE: Using Google and Baidu as example

(1) Search completion time:

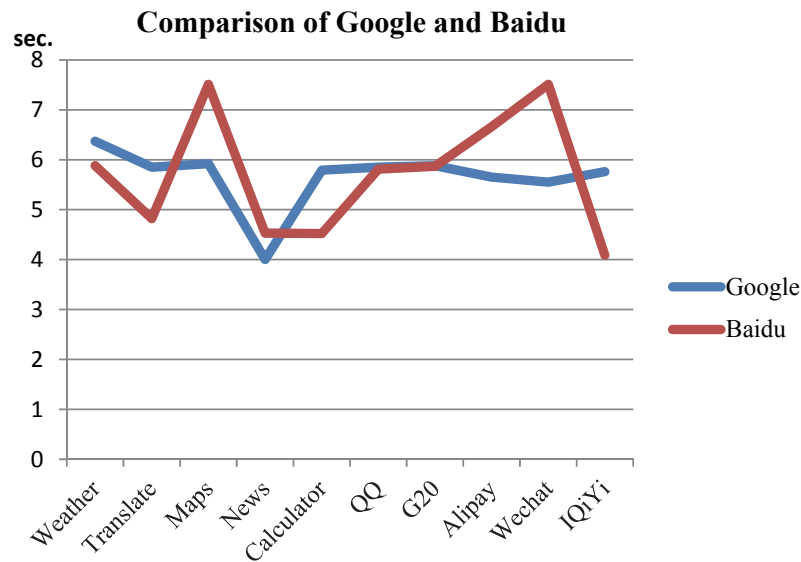
We input each keyword to Google and Baidu SEs and recorded the finish time. The data is shown as Table 3. The comparison of Search completion time is shown as Fig. 3.



**Table 3**

Search completion time in Google and Baidu SEs

Keyword	Google(sec.)	Baidu(sec.)
Weather	6.37	5.88
Translate	5.85	4.82
Maps	5.92	7.51
News	4	4.53
Calculator	5.79	4.52
QQ	5.85	5.81
G20	5.88	5.87
Alipay	5.65	6.66
Wechat	5.55	7.51
IQiYi	5.76	4.09

**Fig. 3.** Comparison of Search completion time in Google and Baidu SEs

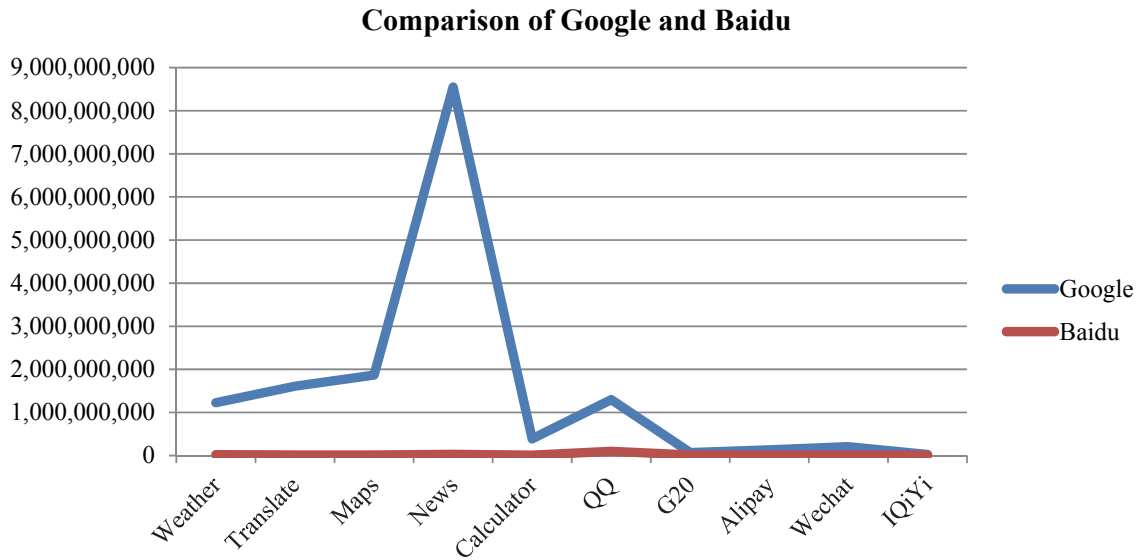
(2) Number of webpages shown in a search task:

We record the number of webpages shown in a search task. The data is shown as Table 4. The comparison of Search completion time is shown as Fig. 4.

**Table 4**

Number of webpages in each keyword search by Google and Baidu SE

Keyword	Google	Baidu
Weather	1,230,000,000	21,100,000
Translate	1,610,000,000	13,600,000
Maps	1,870,000,000	13,100,000
News	8,550,000,000	27,600,000
Calculator	387,000,000	11,400,000
QQ	1,300,000,000	100,000,000
G20	70,000,000	13,300,000
Alipay	132,000,000	11,400,000
Wechat	205,000,000	11,200,000
IQiYi	23,900,000	11,600,000



**Fig. 4.** Comparison of Number of webpages in each keyword search by Google and Baidu SEs

### (3) Precision

We record the number of webpages shown in a search task. We select 50 sites from all of webpages for a sample. And we found the No. of relevant records retrieved and No. of irrelevant records retrieved. And we used the definition of precision to calculate the result of precision. The results of Google and Baidu are shown as Tables 5-6 and the comparison of Google and Baidu is shown as Fig. 5.

**Table 5**

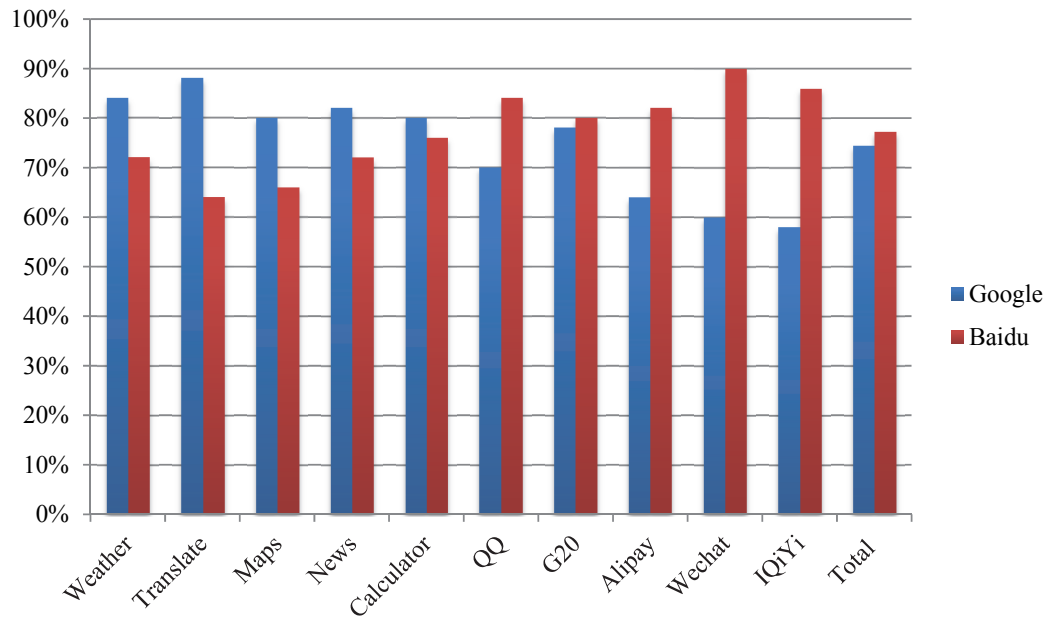
#### Precision of Google SEs

Keyword	Total No. of sites retrieved	Total No. of sites	Relevant	Irrelevant	Precision
Weather	1,230,000,000	50	42	8	84%
Translate	1,610,000,000	50	44	6	88%
Maps	1,870,000,000	50	40	10	80%
News	8,550,000,000	50	41	9	82%
Calculator	387,000,000	50	40	10	80%
QQ	1,300,000,000	50	35	15	70%
G20	70,000,000	50	39	11	78%
Alipay	132,000,000	50	32	18	64%
Wechat	205,000,000	50	30	20	60%
IQiYi	23,900,000	50	29	21	58%
Total	15,377,900,000	500	372	128	74%

**Table 6**

#### Precision of Baidu SEs

Keyword	Total No. of sites retrieved	Total No. of sites evaluated	Relevant	Irrelevant	Precision
Weather	21,100,000	50	36	14	72%
Translate	13,600,000	50	32	18	64%
Maps	13,100,000	50	33	17	66%
News	27,600,000	50	36	14	72%
Calculator	11,400,000	50	38	12	76%
QQ	100,000,000	50	42	8	84%
G20	13,300,000	50	40	10	80%
Alipay	11,400,000	50	41	9	82%
Wechat	11,200,000	50	45	5	90%
IQiYi	11,600,000	50	43	7	86%
Total	234,300,000	500	386	114	77%



**Fig. 5.** Comparison of Precision in Google and Baidu

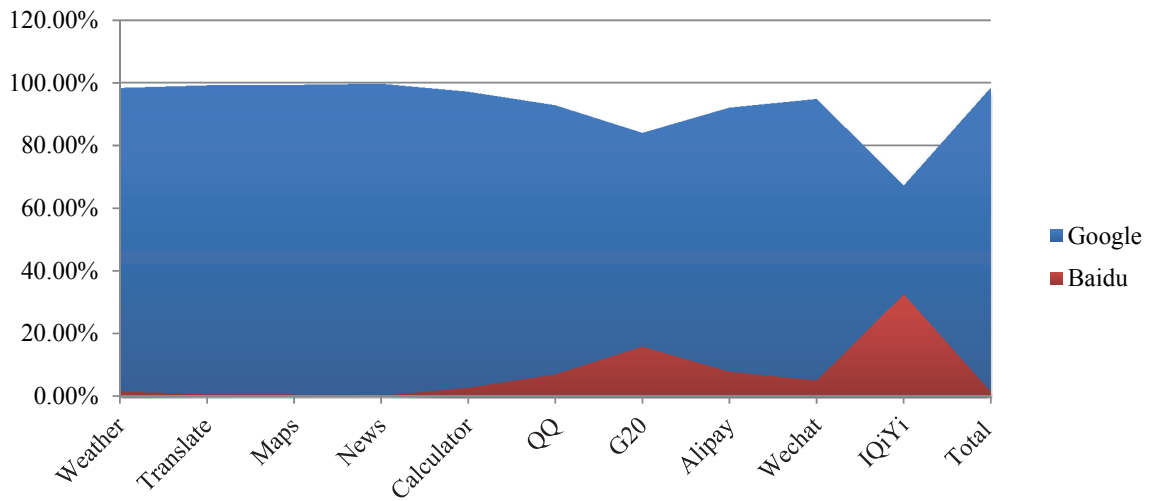
(4) Relative Recall:

We recorded the number of webpages shown in a search task. We used the definition of relative recall to calculate the result of relative recall. The results of Google and Baidu are shown as Table 7 and the comparison of Google and Baidu is shown as Fig. 6.

**Table 7**  
Relative Recall of SEs

Keyword	Google		Baidu	
	Total No. of sites retrieved	Relative recall	Total No. of sites retrieved	Relative recall
Weather	1,230,000,000	98.31%	21,100,000	1.69%
Translate	1,610,000,000	99.16%	13,600,000	0.84%
Maps	1,870,000,000	99.30%	13,100,000	0.70%
News	8,550,000,000	99.68%	27,600,000	0.32%
Calculator	387,000,000	97.14%	11,400,000	2.86%
QQ	1,300,000,000	92.86%	100,000,000	7.14%
G20	70,000,000	84.03%	13,300,000	15.97%
Alipay	132,000,000	92.05%	11,400,000	7.95%
Wechat	205,000,000	94.82%	11,200,000	5.18%
IQiYi	23,900,000	67.32%	11,600,000	32.68%
Total	15,377,900,000	98.50%	234,300,000	1.50%

The present study estimated the four quality features of Google and Baidu SEs, which belong to comprehensive SEs. The result of the study showed that the quality of Google was higher than Baidu. For the number of each keyword searching result, the Google SE was always more than Baidu. Although for the search completion time, Google searched longer than Baidu, but depending on the number of searching results, Google is still better than Baidu. For the precision of each SE, the different keywords lead to a different result. For the relative recall, Google is better than Baidu.



**Fig. 6.** Comparison of Relative Recall in SEs

### 7.2.2 Vertical SE: Using Amazon and JD.COM as example

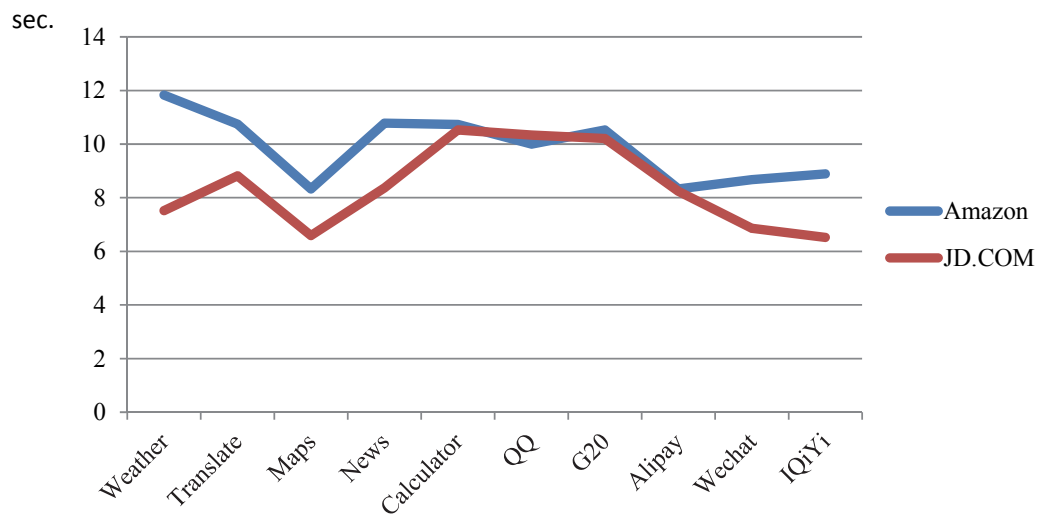
(1) Search completion time:

We input each keyword to Amazon and JD.COM SE and recorded the finish time. The data is shown as Table 8. The comparison of Search completion time is shown as Fig. 7.

**Table 8**

Search completion time in Amazon and JD.COM SEs

Keyword	Amazon(sec.)	JD.COM(sec.)
Weather	11.83	7.52
Translate	10.75	8.82
Maps	8.33	6.59
News	10.78	8.37
Calculator	10.73	10.53
QQ	10	10.34
G20	10.53	10.21
Alipay	8.33	8.24
Wechat	8.68	6.86
IQiYi	8.89	6.52



**Fig. 7.** Comparison of Search completion time in Amazon and JD.COM SEs

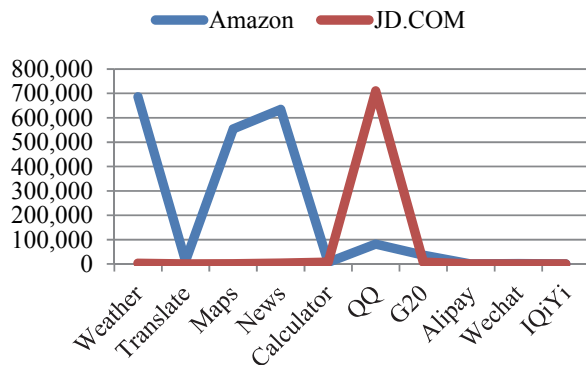
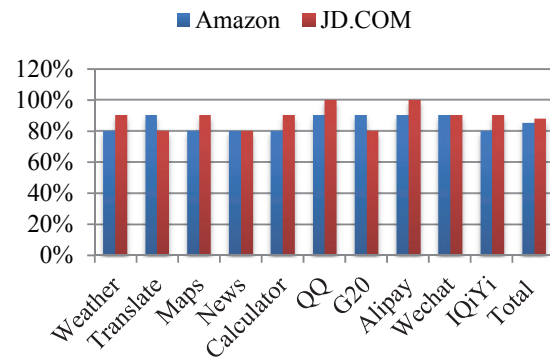
## (2) Number of webpages shown in a search task:

We recorded the number of webpages shown in a search task. The data is shown as Table 9. The comparison of Search completion time is shown as Fig. 8.

**Table 9**

Number of webpages in each keyword search by Amazon and JD.COM SEs

Keyword	Amazon	JD.COM
Weather	686,053	3,500
Translate	12,731	40
Maps	554,234	1,800
News	634,385	4,000
Calculator	7,030	7,400
QQ	81,706	710,000
G20	36,508	7,500
Alipay	14	1
Wechat	2,151	60
IQiYi	17	10

**Fig. 8.** Comparison of Number of webpages in each keyword search by Amazon and JD.COM SEs**Fig. 9.** Comparison of Precision in Amazon and JD.COM

## (3) Precision:

We recorded the number of webpages shown in a search task. We selected 10 sites from all of the webpages for a sample. And we found the No. of relevant records retrieved and No. of irrelevant records retrieved. And we used the definition of precision to calculate the result of precision. The results of Amazon and JD.COM are shown as Table 10-11 and the comparison of Amazon and JD.COM is shown as Fig. 9.

**Table 10**

Precision of Amazon SEs

Keyword	Total No. of sites retrieved	Total No. of sites evaluated	Relevant	Irrelevant	Precision
Weather	686,053	10	8	2	80%
Translate	12,731	10	9	1	90%
Maps	554,234	10	8	2	80%
News	634,385	10	8	2	80%
Calculator	7,030	10	8	2	80%
QQ	81,706	10	9	1	90%
G20	36,508	10	9	1	90%
Alipay	14	10	9	1	90%
Wechat	2,151	10	9	1	90%
IQiYi	17	10	8	2	80%
Total	2,014,829	100	85	15	85%

**Table 11**

## Precision of JD.COM SEs

Keyword	Total No. of sites re-trieved	Total No. of sites evaluated	Relevant	Irrelevant	Precision
Weather	3,500	10	9	1	90%
Translate	40	10	8	2	80%
Maps	1,800	10	9	1	90%
News	4,000	10	8	2	80%
Calculator	7,400	10	9	1	90%
QQ	710,000	10	10	0	100%
G20	7,500	10	8	2	80%
Alipay	1	1	1	1	100%
Wechat	60	10	9	1	90%
IQiYi	10	10	9	1	90%
Total	734,311	91	80	11	88%

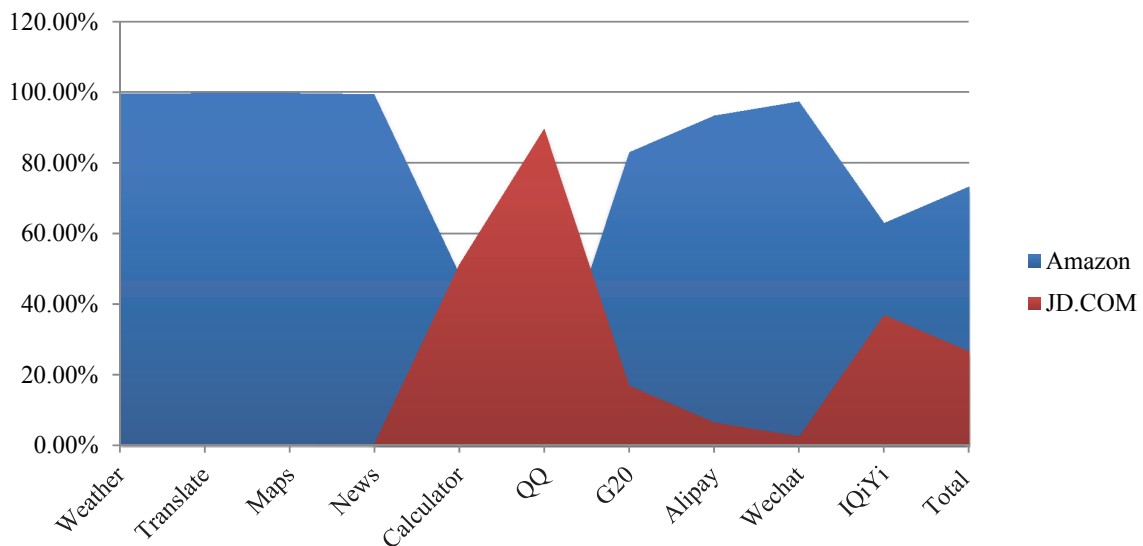
## (4) Relative Recall

We recorded the number of webpages shown in a search task. We used the definition of relative recall to calculate the result of relative recall. The results of Amazon and JD.COM are shown as Table 12 and the comparison of Amazon and JD.COM is shown as Fig. 10.

**Table 12**

## Relative Recall of Amazon and JD.COM SEs

Keyword	Amazon		JD.COM	
	Total No. of sites retrieved	Relative recall	Total No. of sites retrieved	Relative recall
Weather	686,053	99.49%	3,500	0.51%
Translate	12,731	99.69%	40	0.31%
Maps	554,234	99.68%	1,800	0.32%
News	634,385	99.37%	4,000	0.63%
Calculator	7,030	48.72%	7,400	51.28%
QQ	81,706	10.32%	710,000	89.68%
G20	36,508	82.96%	7,500	17.04%
Alipay	14	93.33%	1	6.67%
Wechat	2,151	97.29%	60	2.71%
IQiYi	17	62.96%	10	37.04%
Total	2,014,829	73.29%	734,311	26.71%

**Fig. 10.** Comparison of Relative Recall in Amazon and JD.COM SEs

The present study estimated the four quality features of Amazon and JD.COM SEs, which belong to vertical SEs. The result of the study showed that the quality of Amazon is close to JD.COM. For the number of each keyword search result, the Amazon SE is more than JD.COM in some keywords, but for some keywords, the situation is different. For the search completion time, Amazon searches longer than JD.COM. For the precision of each SE, the different keywords lead to a different result. For the relative recall, Amazon is better than JD.COM.

## 8. Conclusion, Limitation, Future Research Direction

In this study, we have compared the different types of SEs, which are comprehensive SE, and vertical SE. We have selected four common SEs, which are Google, Baidu, Amazon, and JD.COM in the two classifications by researching the four quality features that are Search completion time, Number of webpages shown in a search task, Precision, and Relative Recall. We used ten keywords, compared the quality feature, and found their results. In this research, we found the answer to the two research questions as follows:

For the vertical SE, when we search the keyword, we can get higher precision results, and it does not spend too much time to affirm the results the users needed. But there is a disadvantage that the number of results is less than it was in the comprehensive SEs. For the comprehensive SEs, the responding time is less than the vertical SEs and the number of results is much more than the vertical SEs. And the relative recall of comprehensive SEs is also better than the vertical SEs check of what we need. In a word, before we decide to select which type of SE we should use, we should be sure of our purpose or our requirements.

Although we selected ten keywords, four quality features, four SEs, and two types of SEs to answer the two research questions that are at the beginning of the paper, the research still contains some shortcomings. The limitations of this paper are as follows:

- (1) The data is not very precision data. In this study, we selected ten random keywords and used four SEs to research the four quality features. The data we collected is based on the tools for web developers of Chrome. We did not consider the influences of other factors, such as network speed, location of the researcher, the language of keywords, and so on.
- (2) The categories of the SEs depended on the need of the research project. It was considered convenient to research different kinds of SEs, which ultimately helped determine the conclusions presented in this paper.

In the future, by combining the information in this paper and future research, we hope to improve search effectiveness further. This study researches the parameters of quality, precision, number of results, and relative recall via comparison of information retrieval instruments. These results might lay a foundation for a further study of the users' behavior in search of appropriate information and choosing an adequate SEs to meet the users' needs. In addition, SEs have a long way to go to achieve the quality of the results produced by different requirements.

## References

- Antriksha, S., & Ugrasen S. (2011). Counter measures against evolving search engine Spamming Techniques. *IEEE Conference*, 214-217.
- Blumauer, A., & Hochmeister, M. (2003). Tag-Recommender gestützte Annotation von Web-Dokumenten. *X.media.press Social Semantic Web*, 227-243.
- Burns, E. (2008). Home. Retrieved October 30, 2017, from <https://searchenginewatch.com/sew/study/2066918/almost-billion-us-searches-conducted-july>
- Egele, M., Kolbitsch, C., & Platzer, C. (2009). Removing web spam links from search engine results. *Journal in Computer Virology*, 7(1), 51-62.

- Hochstotter, N., & Koch, M. (2009). Standard parameters for searching behavior in search engines and their empirical evaluation. *Journal of Information Science*, 35(1), 45-65.
- JD.com. (2017). Retrieved October 30, 2017, from <https://en.wikipedia.org/wiki/JD.com>.
- JD.com inks partnership deals with Chinese search engine Sogou. (2017). Retrieved October 30, 2017, from <http://www.thedrum.com/news/2017/10/25/jdcom-inks-partnership-deal-with-chinese-search-engine-sogou>.
- Jiang, M. (2014). Search concentration, bias, and parochialism: A comparative study of Google, Baidu, and Jike's search results from China. *Journal of Communication*, 64(6), 999-1180.
- Jiang, M. (2012). The business and politics of search engines: A comparative study of Baidu and Google's search results of Internet events in China. *New media & Society*, 16(2), 212-233.
- Jopson, B. (2011). Amazon urges California referendum on online tax. Financial Times.
- Meng, C., & Songyun H. (2011). Search engine optimization research for website promotion. *Conference of Information Technology, Computer Engineering and Management Sciences*, 100-103.
- Ramos, A., & Cota, S. (2004). Insider guide to SEO: How to get your Website to the top of the search engines. *Jain Publishing Company*, Fremont, CA.
- Robinson, M.L., & Wusteman, J. (2007). Putting Google Scholar to the test: a preliminary study. *Program: Electronic Library and Information Systems*, 41(1), 71-80.
- Sterling, G. (2012). Bing and Google gain market share while Yahoo drops, available at: <http://searchengineland.com/bing-and-google-gain-market-share-while-yahoo-drops-114140>.
- Tauqeer, A., U. (2012). A comparative study of Google and Bing search engines in context of precision and relative recall parameter. *International Journal on Computer Science and Engineering (IJCSE)*, 4(1).
- Termehchy, A., & Winslett, M. (2009). Generic and effective semi-structured keyword search. *Proceedings of the First International Workshop on Keyword Search on Structured Data – KEYS*.
- "The 100 most popular Google keywords". (2017). Available at: <https://www.siegemedia.com/seo/most-popular-keywords>.
- "Top Baidu Searches 2016". (2016). Available at: <https://www.chinaskinny.com/blog/top-baidu-searches-2016/>
- Uyar, A. (2009). Investigation of the accuracy of search engine hit counts. *Journal of Information Science*, 35(4), 469-80.
- Visser, E.B., & Weideman, M. (2011). An empirical study on website usability elements and how they affect search engine optimization. *South African Journal of Information Management*, 13(1), 1-9.
- Weideman, M. (2004). Empirical evaluation of one of the relationships between the user, search engines, metadata and websites in three-letter .com websites. *South African Journal of Information Management*, 6(3).
- Wu K., Jin H., Zheng R., & Zhang Q. (2010). A vertical search engine based on visual and textual features. Entertainment for Education. Digital Techniques and Systems. Edutainment 2010. *Lecture Notes in Computer Science*, 6249.
- Ya, X., & David, M. (2009). Evaluating web search using task completion time. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, July 19-23, 2009, Boston, MA, USA.
- Zhang, J., & Dimitroff, A. (2005). The impact of webpage content characteristics on webpage visibility in search engine results (part I). *Information Processing and Management*, 41(3), 665-690.
- Ziyang, L. (2011). Enhancing the usability of complex structured data by supporting keyword searches. *Arizona State University Tempe, AZ, USA ©2011*.

