

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

available at [www.sciencedirect.com](http://www.sciencedirect.com)journal homepage: [www.elsevier.com/locate/iilr](http://www.elsevier.com/locate/iilr)

# Search engines' responses to several search feature selections

Jin Zhang<sup>a,b,\*</sup>, Wei Fei<sup>c</sup>

<sup>a</sup> International Collaborative Academy of Library and Information Science, Wuhan University, Wuhan, Hubei 430072, PR China

<sup>b</sup> School of information Studies, University of Wisconsin Milwaukee, 3210N Maryland Ave, Bolton Hall 532, Milwaukee, WI 53211, USA

<sup>c</sup> School of Information Management, Wuhan University, Wuhan, Hubei, 430072, PR China

## KEYWORDS

Search engines;  
Evaluation;  
Search feature;  
Retrieval effectiveness;  
Regression analysis;  
Information retrieval

**Abstract** The authors of this paper investigated the impact of the advanced search features of three common search engines on retrieval result performance: Yahoo, Google, and Live Search. The authors analyzed 240 search queries with different information need emphases to determine retrieval effectiveness differences among regular search, title search, exact phrase search, and PDF file format restriction search. A one-way ANOVA method and regression analysis method were used for the study. It was found that the PDF file format restriction search achieved the best retrieval performance among Yahoo, Google and Live Search. The regular search achieved the best web page ranking performance among Yahoo, Google, and Live Search. The findings of this study can be used to assist users in formulating an appropriate search strategy to improve search effectiveness, and to shed light on how search engines react to different types of search features in terms of retrieval effectiveness.

© 2010 Elsevier Ltd. All rights reserved.

## Introduction

The Internet has become a global data communication system and information resource, changing the way people look for information. It has become a primary means to search information due to its richness, coverage, diversity,

and dynamics, plus powerful search engines. Its impact on society is significant and unprecedented. The Internet has brought new forms of social interactions, networks, and online activities because of its accessibility and availability. It was reported that the number of unique users reached a new record high: One billion in December 2008 (Wallop, 2009). However, the actual number of Internet users might be even higher.

An Internet search engine is a key tool designed to identify information on the Internet. One survey shows 89% of Internet users use search engines to find information, second only to the 91% who send or read e-mail (Pew Research Center, 2008). A search engine responds to a query submitted from a user with a returned results list

\* Corresponding author. School of information Studies, University of Wisconsin Milwaukee, 3210N Maryland Ave, Bolton Hall 532, Milwaukee, WI 53211, USA. Tel.: +1 414 229 2712; fax: +1 414 229 6699.

E-mail addresses: [jzhang@uwm.edu](mailto:jzhang@uwm.edu) (J. Zhang), [fei.wayne@gmail.com](mailto:fei.wayne@gmail.com) (W. Fei).

from billions and billions of indexed web pages. The list is ranked by the relevance of each result to the query.

A search engine provides users with a variety of search features such as: exact phrase search, Boolean search, language restriction, file type restriction, title search, and domain restriction. These features are supposed to facilitate users' searches, meet various special types of information needs, and improve search effectiveness. Some search features are commonly provided by all the search engines while some advanced search features may vary in different search engines. A study suggests that a search engine should provide users with powerful advanced features to support query processes and achieve effective searches (Wu & Chuang, 2008).

The evaluation of search engines ranges across many fronts, beginning with search engine evaluation from users' perspective (Crudge & Johnson, 2004; Spink, 2002; Su, 2003), to crossing language search (Zhang & Lin, 2007), to image search (Hassan & Zhang, 2001), to meta-search engine (Zhang & Chueng, 2003), to e-commerce search (Jansen & Molina, 2006), to multimedia search (Ozmutlu, Spink, & Ozmutlu, 2003), to search engine bias (Mowshowiz & Kawaguchi, 2005; Vaughan & Thelwall, 2004), to search engine result overlapping (Spink, Jansen, Blakely, & Koshman, 2006), to search engine result ranking (Bar-Ilan, 2006; Bar-Ilan, Keenoy, Yaari, & Levene, 2007), to research engine result consistency (Thelwall, 2008), and to sponsored web pages in search results (Nicholson et al., 2006). Two advanced search features (exact title search and URL search) were selected to examine the impact of non-topical terms and semi-topical terms on query expansion and one study showed that the search results would improve if the queries are restricted to the exact title search or URL search (Fattahi, Wilson, & Cole, 2008). Contrary to expectation, in another study Boolean search strategies delivered negative retrieval effectiveness (Ford, Miller, & Moss, 2002). According to Eastman and Jansen (2003), the use of most query operators such as AND, OR, MUST APPEAR (C), or PHRASE (" ") were not used by the majority of searchers and this had no significant effect on coverage, relative precision, or search results ranking.

The above analysis shows that although a lot of studies on search engine evaluation have been done, each of them has had its own emphasis. None of them evaluated and compared search features across multiple search engines and evaluation on search features, like PDF file format restriction search, was not addressed.

It is widely recognized that users' information needs are sophisticated and dynamic. Users' information needs vary in different situations and contexts. An information retrieval system should be able to provide its users with various search features and means to cope with complicated users' information needs. There is no exception for Internet search engines. Internet search engines like Google and Yahoo are equipped with an array of search features to satisfy users' information needs. The importance of a study which investigates search features of the search engines is to enable users to better understand the characteristics of the search features in the major search engines and to better meet their information needs by selecting appropriate search features.

Use of an advanced search feature in a search engine certainly impacts the returned results list from a search engine. Users may ask whether the use of an advanced search feature would make the returned web pages more relevant and accurate, and the ranking of the returned web pages in the list more accurate. Unfortunately, the answers to these questions are rarely found in the literature.

The main objective of this study is to investigate whether the use of advanced search features would improve the effectiveness and accuracy of search engine retrieval results through an experimental study. The independent variable and dependent variable in this study are search feature and retrieval effectiveness in a search engine, respectively. The findings of this study can be used to help users select an appropriate search engine feature to improve search effectiveness, and to gain insight into how a search engine responds to different types of search features in terms of retrieval result effectiveness.

## Research method description

A search engine commonly provides users with a group of advanced search features in addition to simple keyword search. The advanced search features vary in different search engines. They may include title search, exact phrase search, Boolean search, language restriction, file type restriction, country restriction, domain restriction, URL search, meta word search, date restriction, similar pages, linked pages, returned result number restriction, usage rights restriction, numeric range, and safe search filter. Advanced search offers a wide spectrum of search options for searchers. These search features are designed to meet users' various special information needs and therefore users should achieve better retrieval performance.

The overarching research question for this study is whether the use of advanced search features enhances retrieval effectiveness in a search engine. Based on the research question three null hypotheses of the study were developed and are stated as follows:

- *H1: There are no significant differences among the searches without the advanced search features and searches with the advanced search features in terms of retrieval effectiveness in Yahoo.*
- *H2: There are no significant differences among the searches without the advanced search features and searches with the advanced search features in terms of retrieval effectiveness in Google.*
- *H3: There are no significant differences among the searches without the advanced search features and searches with the advanced search features in terms of retrieval effectiveness in Live Search.*

Hypotheses H1, H2 and H3 were used to test retrieval performance in each of the three search engines. The independent variable was advanced search feature used in a search query and the dependent variable was retrieval effectiveness of a search engine search. The control group consisted of the searches without a search feature and the treatment group consisted of the searches with a search feature.

The measurement for retrieval performance in this study was revised precision. The traditional precision for a search is defined as the ratio of the number of the retrieved relevant documents in a result set to the size of the result set. The relevance judgment result for a retrieved document is binary. In other words, a document is regarded as either relevant or irrelevant. In fact, a retrieved document may fall into a grey area between relevance and irrelevance. The definition of the traditional precision does not fit the evaluation of search engine retrieval results. That is because a search engine usually responds to a regular query with a lengthy results list. The size of the results list can easily reach thousands, even millions. Relevance judgment of a web page to a query is time consuming. It was not feasible for the researchers to examine the relevance of all the returned web pages of a query. In addition, it was reported that most users only browse the web pages in first page of a results list (Jacsó, 2008; Jansen & Spink, 2006; Kumar, Suri, & Chauhan, 2005). Leighton and Srivastava (1999) argue that the first three results are very important, and then the next seven, and then the next ten. Users usually ignore and skip the rest of the web pages in the results list. Due to these reasons, a revised precision was introduced to evaluate search effectiveness in this study.

First, the relevance of a retrieved web page to a query was evaluated by the researchers and a five-relevance-level scale was applied in Fig. 1. After the title and full-text of a web page are browsed analyzed, a relevance score was given to the web page. The defined relevance scale ranges from 0 to 1. For instance, if the relevance of a web page to a query was in between relevant and the most relevant, then 0.75 as its relevance score may have been assigned. If, of two web pages one was not relevant and the other was fully relevant, then they were assigned 0 and 1 respectively. If a returned web page in the results list corresponded to a dead link, it was skipped. In other words, the dead links in a results list were not considered in the study.

During the relevance judgment process, the investigators examined all paragraphs, figures, and tables in each of the retrieved web pages returned from a search engine. If all the paragraphs, figures, and tables in the web page were related to a query, the relevance score for the web page was 1. If half of the paragraphs, figures, and tables in the web page were related to the query, 0.5 was assigned to the web page as its relevance score. If none of the paragraphs, figures, and tables in the web page was related to the query, the web page received 0 as its relevance score.

The relevance judgment for web pages was conducted by the researchers of this study.

It is widely recognized that the relevance judgment for a retrieved document/web page is subjective. Many factors, such as search domain background, understanding of the information need, retrieval expertise, and retrieval result presentation format, play a role in the relevance judgment. In this study, the investigators of this study were responsible for the relevance judgment for the retrieved web pages from the search engines. The positive side of this

method is that the investigators developed the queries, understood the information needs behind these queries, and applied consistently the relevance evaluation criteria to the retrieved web pages. On the other hand, since the relevance judgment task in this study was not conducted by users, the results of the study may be different from the results performed by users.

The number ( $n$ ) of the examined web pages in a results list was determined. In our case the parameter  $n$  is equal to 100 for all the retrieval results lists. That is, only the top 100 web pages in a search results list were examined and those web pages were judged accordingly.

The revised precision (search effectiveness measure) is defined as:

$$P(w_i) = \frac{\sum_{i=1}^n R_i(w_i)}{n} \quad (1)$$

in Equation (1),  $P(w_i)$  is the revised precision,  $R_i(w_i)$  is the relevance score of a web page ( $w_i$ ) in a retrieval results list, and  $n$  is the number of all the examined web pages in a results list from a search engine.

The three search engines selected for the experimental study – Google, Yahoo, and Live Search – have dominated the search engine use market (Sullivan, 2006). Google, Yahoo and MSN (Live Search) represent 85.3% of searches in the US. The search engines have been widely recognized as primary search engines. Each search engine provides users with rich advanced search features.

The following three search features in the three search engines were identified and selected for the study:

- Title search (search terms only appear in a web page title)
- PDF file format restriction search (the retrieved web pages' file format is PDF)
- Exact phrase matching search (phrase in a query is exactly matched with the phrase in retrieved web pages)

These three search features were identified and selected for the study for specific reasons. It is, for example, widely recognized that keywords in a title are more important than keywords in full-text in terms of describing subject content (Fattahi et al., 2008; Zhang & Dimitroff, 2005). It is natural for users to consider the title search feature as the first option in their searches. The PDF file format restriction search returns a group of the documents which usually are published journal papers, technical reports, conference papers, or other types of formal documents. These returned results are more important for researchers if they look for formal documents such as research papers for research. Exact phrase search allows searchers to search for a phrase or short sentence in the retrieved web pages. In this case, keywords in a phrase are no longer separated as individual query terms. Instead a phrase/sentence is treated as a complete fixed term. For this reason, exact phrase search may be good for a specific search. In addition, these three search features are all available in the three investigated search engines. Consequently, retrieval performance of these three search features can be compared across the investigated search engines. A regular search without any advanced

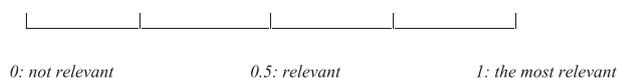


Figure 1 Relevance scale display.

**Table 1** Display of the search queries in the study.

Medicine and health	Culture and education	Information and technology	Economy and business
Obesity and heart disease	World cultural diversity	Information retrieval technology	Global financial crisis 2008
Diet and diabetes	Distance education	Information management	Economic stimulus
Sports and health	Information literacy	Information security	Industrial pollution
Mental health	World cultural heritage	Intellectual property	Food security
Keep fit	Folk customs	Electronic commerce	International trade and marketing

search feature is also included in the study. The results from a regular search can serve as a baseline for comparison and analysis in a search engine.

In order to achieve a sound and plausible experimental result, 20 search tasks were generated for the study, and the contents of these 20 tasks focused on four different areas: medicine and health; culture and education; information and technology; and economy and business. In other words, each of these four areas corresponded to 5 tasks. The 20 search tasks were used in the three search engines respectively. Four queries were formulated based on each of the 20 search tasks. As a result, 80 queries were yielded and submitted to each of the three search engines. Among the 80 queries, 20 queries were used for title search, 20 queries for exact phrase search, 20 queries for PDF format filter search, and 20 queries for regular search.

The 20 search queries, which are categorized into the 4 groups, are listed in Table 1.

After a query was submitted to a search engine, the search engine returned a retrieval results list. The top 100 returned web pages were examined, the relevance of each web page in the returned web page set was judged, the relevance judgment results were tabulated, and retrieval effectiveness of each search was calculated using Equation (1) for further data analysis. Sponsored web pages in a results list were excluded for study. In addition, if a returned hyperlink was a dead link in a results list, it was not considered.

All collected resultant retrieval effectiveness data were categorized into four groups: regular search group; title search group; PDF format group; and phrase exact matching group. A one-way analysis of variance (ANOVA) method was applied to the four groups of data to test whether there were significant differences among the four groups. If there were significant differences among the four groups, The Tukey analysis method was used to detect what caused the significant differences among the four groups. A one-way ANOVA method was applied to the data from the three search engines respectively so that the advanced search

feature behaviors in the three different search engines could be revealed and demonstrated separately. In addition, the resultant data from the three search engines were merged into one data set and a one-way ANOVA method was also applied to the merged and integrated data set so that the advanced search feature behaviors in the three different search engines as a whole could be discovered and illustrated. The ANOVA results indicated whether the proposed null hypotheses were rejected or accepted.

As we know, a search engine returns a retrieval results list in response to a query. The web pages in the results list are not randomly presented to users. Instead, they are ranked based on their similarities to the search query. It is widely recognized that users tend to read and browse the web pages at the top of a retrieval results list from a search engine. If the web pages with high relevance scores are located at bottom of the list, the chance that the web pages are browsed is slim. Therefore, web page ranking should be considered as one of the important factors for search engine retrieval effectiveness evaluation. The regression analysis method was used to investigate the retrieval result ranking change patterns of the web pages in the ranking results lists in the three search engines. The regression equations were identified, the regression curves were produced, and the goodness of fit for each regression analysis was provided. The regression analysis method also allows readers to understand the degree to which the retrieval effectiveness values in a results list fit into the identified patterns/models. The regression analysis method added a unique dimension to retrieval effectiveness analysis of search engines.

## Results and discussion

In this study, 240 searches were submitted to three search engines (*Yahoo*, *Google*, and *Live Search*). A total of 24,000 web pages were examined and their relevance judgments were made accordingly. Retrieval effectiveness of each

**Table 2** Descriptive summary of the Yahoo data.

Variable	N	Mean	Std. deviation	Std. error	95% Confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
Ti	20	0.6910	0.03262	0.00729	0.6757	0.7063	0.65	0.75
Re	20	0.6157	0.03697	0.00827	0.5984	0.6330	0.54	0.68
Ph	20	0.5242	0.03863	0.00864	0.5061	0.5423	0.43	0.58
PDF	20	0.7688	0.04060	0.00908	0.7498	0.7878	0.67	0.83
Total	80	0.6499	0.09820	0.01098	0.6281	0.6718	0.43	0.83

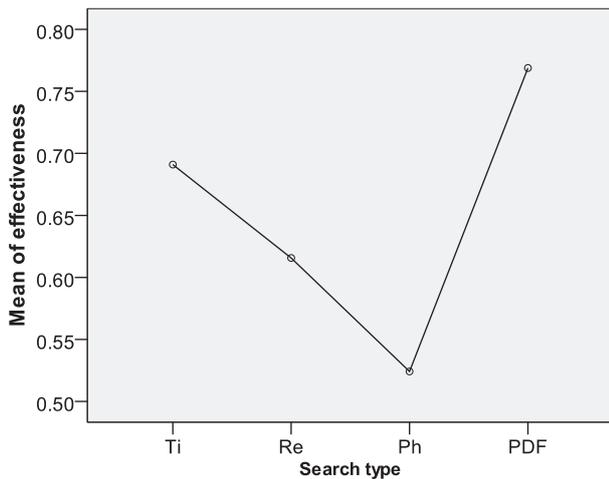


Figure 2 Mean plots for the Yahoo test.

search was calculated. The data analysis tool was the SPSS statistics analysis software package (Version 16). The significance level ( $p$ ) or  $sig$  was 0.05 for all the tests in this study. In other words, if  $p$  or the  $sig$  was smaller than 0.05 in a test, the finding of the test was statistically significant and the corresponding null hypothesis was rejected.

### Data results and data analysis for Yahoo

The one-way ANOVA test results illustrate that the degree of freedom (DF) was 3, the  $F$  value was 156.958, the mean square was 0.219, and the  $p$ -value was 0.000, which is smaller than 0.05. Therefore, the null hypothesis  $H_1$  was rejected. This means that there were significant differences among search results without advanced search features and search results with advanced search features in terms of retrieval effectiveness in Yahoo.

Table 2 shows the descriptive summary of the Yahoo resultant data. In the first column "Variable", Ti, Re, Ph, and PDF stand for title search, regular search (without using any search feature), exact phrase search, and PDF file

format restriction search, respectively. The mean plots are displayed in Fig. 2 where the X-axis and Y-axis represent the search feature and retrieval effectiveness mean respectively. The retrieval effectiveness mean of the exact phrase search (0.5242) was the smallest among the four groups. It was even smaller than the effectiveness of the baseline, the regular search without using any search feature. The greater a retrieval effectiveness mean was, the better the retrieval effectiveness, and vice versa. The smallest standard deviation of the title search (0.03262) indicates that title search effectiveness was the most stable compared to other types of search. The greatest standard deviation was 0.04060 (The PDF file restriction search). The lower boundary and upper boundary for the PDF file format restriction search were 0.7498 and 0.7878 at 95% confidence interval, respectively. They were the highest scores in the lower boundary and upper boundary categories, respectively. The higher the score of the lower (upper) boundary, the better the retrieval performance was.

The minimum and maximum means for the PDF file format restriction search were 0.67 and 0.83, respectively. They were the highest scores in the minimum and maximum mean categories respectively. On the other hand, the lower boundary and upper boundary for the exact phrase search were 0.5061 and 0.5423 at 95% confidence interval, respectively. They were the lowest scores in the lower boundary and upper boundary categories respectively. The minimum and maximum means for the exact phrase search were 0.43 and 0.58, respectively. They were the lowest scores in the minimum and maximum mean categories, respectively. Other resultant data like standard deviation error are also listed in Table 2.

Since the null hypothesis  $H_1$  was rejected in this case due to the low  $p$ -value (0.000), a Tukey HSD method was carried out to further determine the reasons for the rejection. The generated Tukey HSD analysis results are presented in Tables 3 and 4. The definitions of the factors ( $I$  and  $J$ ) Ti, Re, Ph, and PDF in Tables 3 and 4 are the same as these in Table 2. A positive value of  $I-J$  indicates that the factor  $I$  outperformed the factor  $J$ . The sign "\*\*" in the second column in Table 3 indicates that the corresponding mean difference was significant at the 0.05 level. The mean difference ( $I-J$ )

Table 3 Multiple comparisons of the Tukey HSD for the Yahoo test.

(I) factor	(J) factor	Mean difference ( $I - J$ )	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Ti	Re	0.07530*	0.01180	0.000	0.0443	0.1063
	Ph	0.16680*	0.01180	0.000	0.1358	0.1978
	PDF	-0.07780*	0.01180	0.000	-0.1088	-0.0468
Re	Ti	-0.07530*	0.01180	0.000	-0.1063	-0.0443
	Ph	0.09150*	0.01180	0.000	0.0605	0.1225
	PDF	-0.15310*	0.01180	0.000	-0.1841	-0.1221
Ph	Ti	-0.16680*	0.01180	0.000	-0.1978	-0.1358
	Re	-0.09150*	0.01180	0.000	-0.1225	-0.0605
	PDF	-0.24460*	0.01180	0.000	-0.2756	-0.2136
PDF	Ti	0.07780*	0.01180	0.000	0.0468	0.1088
	Re	0.15310*	0.01180	0.000	0.1221	0.1841
	Ph	0.24460*	0.01180	0.000	0.2136	0.2756

\*The mean difference is significant at the 0.05 level.

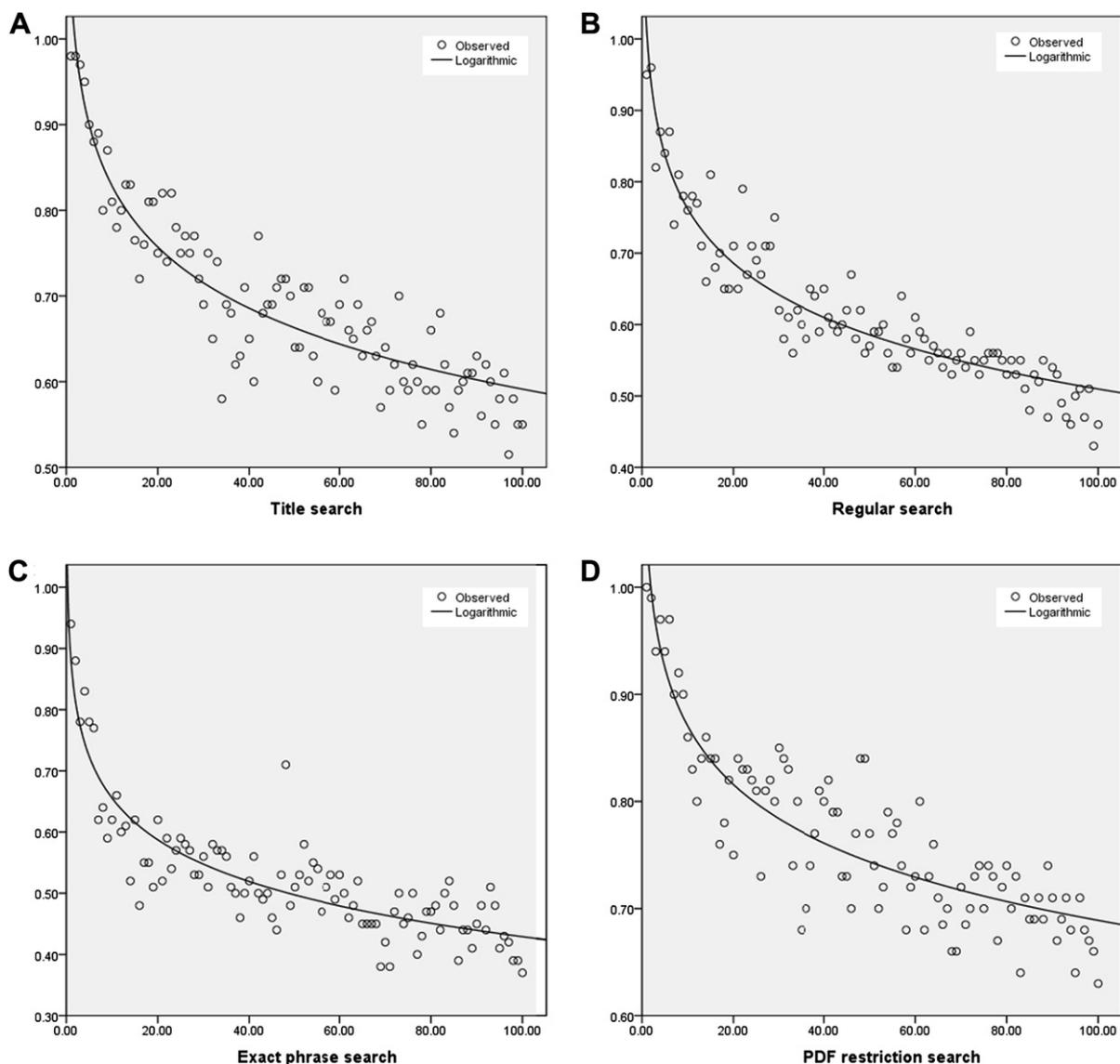
**Table 4** Homogeneous subsets for the Yahoo test.

Factor	N	Subset for alpha = 0.05			
		1	2	3	4
Ph	20	0.5242			
Re	20		0.6157		
Ti	20			0.6910	
PDF	20				0.7688
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

column in Table 3 demonstrates that all the mean differences between the PDF file restriction search and other types of search were positive and significant: 0.07780 (the mean of PDF file format restriction search – the mean of title search), 0.15310 (the mean of PDF file format restriction search – the mean of regular search), and 0.24460 (the

mean of PDF file format restriction search – the mean of exact phrase search). In other words, the PDF file format restriction search in this study outperformed the title search, the regular search, and the exact phrase search as well. The title search is second to the PDF file format restriction search because the mean difference between the



**Figure 3** Regression analysis for four search types in Yahoo.

**Table 5** Descriptive Summary of the Google data.

Variable	N	Mean	Std. deviation	Std. error	95% Confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
Ti	20	0.6414	0.03071	0.00687	0.6271	0.6558	0.59	0.70
Re	20	0.5797	0.03000	0.00671	0.5657	0.5937	0.53	0.63
Ph	20	0.5062	0.04272	0.00955	0.4862	0.5262	0.40	0.58
PDF	20	0.6735	0.03692	0.00826	0.6562	0.6908	0.64	0.76
Total	80	0.6002	0.07311	0.00817	0.5839	0.6165	0.40	0.76

title search and the regular search was 0.07530, and the mean difference between the title search and exact phrase search was 0.16680. The regular search was in the third place because the mean difference between the regular search and the exact phrase search was 0.09150. Performance of the exact phrase search was the worst among the four search types. All the *sig* values are 0.000 in Table 3.

In a *Tukey HSD* analysis, if the factors are grouped into a subset, these factors are not significantly different. In the test four separate homogeneous subsets were produced in Table 4. That is, each search type corresponded to a homogeneous subset. No two search types formed a homogeneous subset. It confirms that there were significant differences among the four search types in terms of retrieval effectiveness.

In order to examine and observe the relevance ranking distribution of retrieved web pages in a retrieval results list, the average relevance scores of all searches in one search type had to be calculated and then a regression analysis conducted for each of the four search types. In Fig. 3, the *X*-axis and *Y*-axis are defined as the position of a web page in a retrieval results list and the relevance score of a web page, respectively. A greater *R* square value indicates a better regression curve fit. In other words, a greater *R* square score is associated with a better web page ranking result in a retrieval results list.

The regression analysis results for the title search were as follows: The *R* square was 0.838, *F* value was 508.572, and *Sig* was 0.000. The regression equation was a logarithmic model ( $b_0 = 1.066$ ,  $b_1 = -0.103$ ). The regression curve is shown in Fig. 3A. The standard logarithmic model/equation is defined in Equation (2).

$$Y = b_0 + b_1 \times \ln(X) \quad (2)$$

The regression analysis results for the regular search were as follows: The *R* square was 0.882, *F* value was 730.472, and *Sig* was 0.000. The regression equation was a logarithmic model ( $b_0 = 1.014$ ,  $b_1 = -0.110$ ). The regression curve is shown in Fig. 3B.

The regression analysis results for the exact phrase search were as follows: The *R* square was 0.781, *F* value was 348.472, and *Sig* was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.882$ ,  $b_1 = -0.098$ ). The regression curve is shown in Fig. 3C.

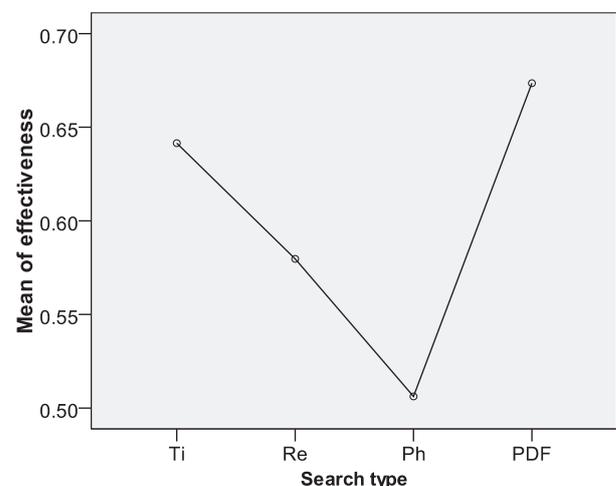
The regression analysis results for the *PDF* file restriction search were as follows: The *R* square was 0.788, *F* value was 363.475, and *Sig* was 0.000. The regression equation was a logarithmic model ( $b_0 = 1.052$ ,  $b_1 = -0.079$ ). The regression curve is shown in Fig. 3D.

It is apparent that the regular search achieved the best regression curve fit due to its higher *R* square value (0.882). Fig. 3B illustrates that the web pages are nicely scattered around the generated regression equation curve. The exact phrase search achieved the worst because of its lower *R* square (0.781).

### Data results and data analysis for Google

The one-way *ANOVA* test results illustrate that *DF* was 3, the *F* value was 86.558, the mean square was 0.109, and the *p*-value/*sig* was 0.000, which is smaller than 0.05. Therefore, the null hypothesis *H*<sub>2</sub> was rejected. This means that there were statistical differences among searches without advanced search feature and searches with advanced search features in terms of retrieval effectiveness in Google.

Table 5 shows the descriptive summary of the Google resultant data. In the first column "Variable", the definitions of Ti, Re, Ph, and PDF are the same as Table 2. The Google test mean plots are presented in Fig. 4 where the definitions of the *X*-axis and *Y*-axis are the same as those in Fig. 2. The retrieval effectiveness mean of the exact phrase search (0.5062) was the smallest while that of the *PDF* file format restriction search (0.6735) was the greatest. The smallest standard deviation of the regular search (0.03000) indicates that the regular search effectiveness was the most stable compared to other types of searches. The greatest standard deviation was 0.04272 (the exact phrase search). The lower boundary and upper

**Figure 4** Mean plots for the Google test.

**Table 6** Multiple comparisons in Tukey HSD for the Google test.

(I) Factor	(J) Factor	Mean difference (I – J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Ti	Re	0.06175*	0.01122	0.000	0.0323	0.0912
	Ph	0.13525*	0.01122	0.000	0.1058	0.1647
	PDF	-0.03205*	0.01122	0.028	-0.0615	-0.0026
Re	Ti	-0.06175*	0.01122	0.000	-0.0912	-0.0323
	Ph	0.07350*	0.01122	0.000	0.0440	0.1030
	PDF	-0.09380*	0.01122	0.000	-0.1233	-0.0643
Ph	Ti	-0.13525*	0.01122	0.000	-0.1647	-0.1058
	Re	-0.07350*	0.01122	0.000	-0.1030	-0.0440
	PDF	-0.16730*	0.01122	0.000	-0.1968	-0.1378
PDF	Ti	0.03205*	0.01122	0.028	0.0026	0.0615
	Re	0.09380*	0.01122	0.000	0.0643	0.1233
	Ph	0.16730*	0.01122	0.000	0.1378	0.1968

\*The mean difference is significant at the 0.05 level.

boundary for the PDF file format restriction search was 0.6562 and 0.6908 at 95% confidence interval, respectively. They were the highest scores in the lower boundary and upper boundary categories, respectively. The minimum and maximum means for the PDF file format restriction search were 0.64 and 0.76, respectively. They were also the highest scores in the minimum and maximum mean categories. On the other hand, observe that the lower boundary and upper boundary for the exact phrase search were 0.4862 and 0.5262 at 95% confidence interval, respectively. They were the lowest scores in the lower boundary and upper boundary categories, respectively. The minimum and maximum means for the exact phrase search were 0.40 and 0.58, respectively. They were also the lowest scores in the minimum and maximum mean categories.

Since the null hypothesis H2 was rejected in this case due to the low p-value (0.000), a Tukey HSD method was used to further determine the reasons for the rejection. The produced Tukey HSD analysis results are presented in Tables 6 and 7. The definitions of the factors (I and J) Ti, Re, Ph, and PDF in Tables 6 and 7 are the same as those in Table 5. The mean difference (I–J) column in Table 6 shows that all the mean differences between the PDF file restriction search and other types of search were positive and significant: 0.03205 (the mean of PDF file format restriction search - the mean of title search), 0.09380 (the mean of the PDF file format restriction search – the mean of regular search), and 0.16730 (the

mean of the PDF file format restriction search – the mean of exact phrase search). In other words, the PDF file format restriction search achieved the best retrieval performance among the four groups. The title search is in second place because the significant mean difference between the title search and the regular search was 0.06175, and the significant mean difference between the title search and exact phrase search was 0.13525. The regular search is in third place because the significant mean difference between the regular search and the exact phrase search was 0.07350. Performance of the exact phrase search was the worst among the four search groups. All the sig values for the mean differences were 0.000 except for the mean difference between the title search and the PDF file format restriction search (0.028) in Table 6.

In the Tukey HSD analysis, four separate homogeneous subsets were produced in Table 7. That is, each search type corresponds to a homogeneous subset, which is consistent with the finding that there were significant differences among the four search types in terms of retrieval effectiveness.

A regression analysis was carried out for each of the four search types. In Fig. 5, the definitions of the X-axis and Y-axis are the same as those in Fig. 3.

The regression analysis results for the title search were as follows: The R square was 0.845, the F value was 730.472, and Sig was 0.000. The regression equation was

**Table 7** Homogeneous subsets for the Google test.

Factor	N	Subset for alpha = 0.05			
		1	2	3	4
Ph	20	0.5062			
Re	20		0.5797		
Ti	20			0.6414	
PDF	20				0.6735
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

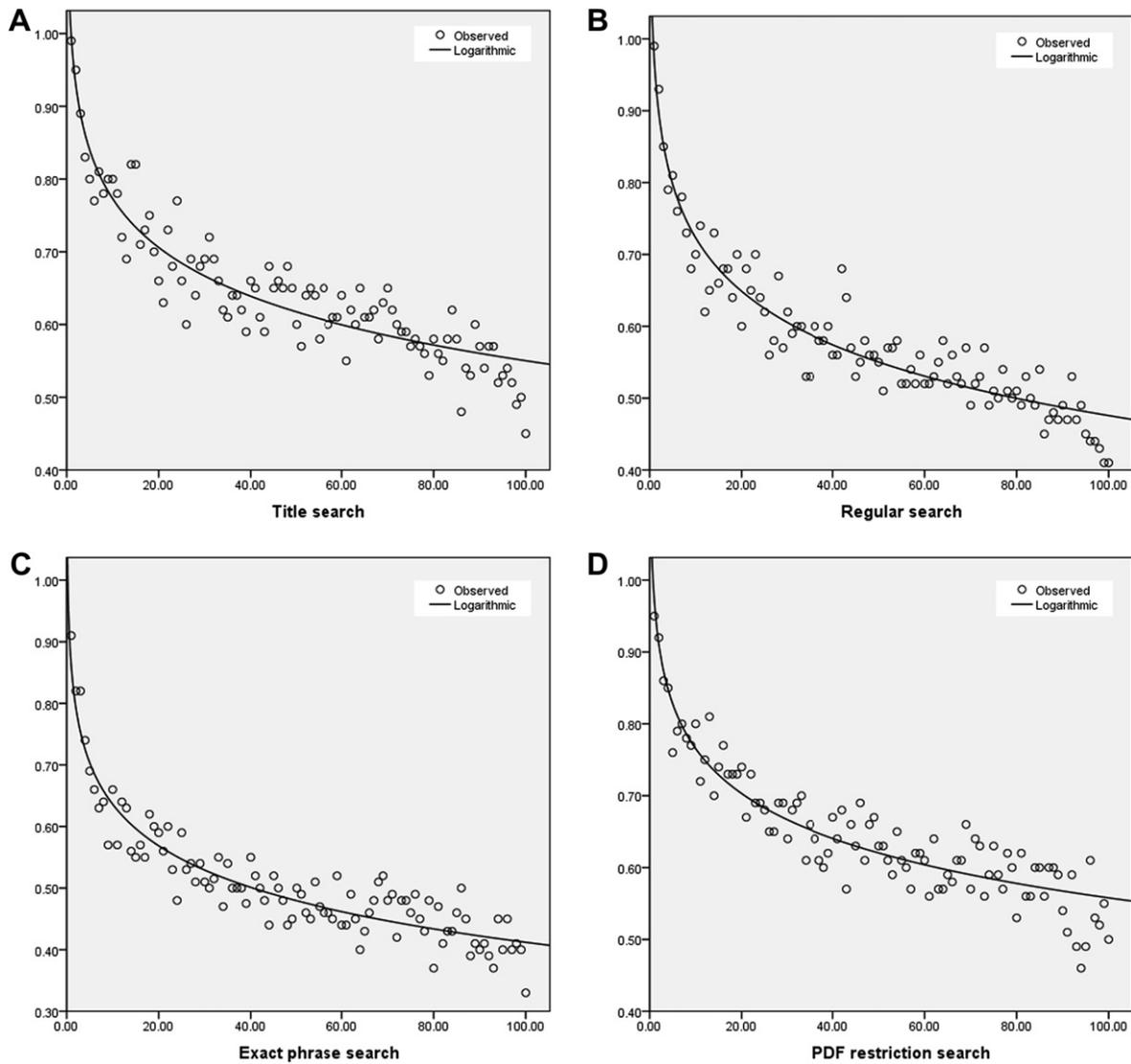


Figure 5 Regression analysis for four search types in Google.

a logarithmic model ( $b_0 = 0.996$ ,  $b_1 = -0.097$ ). The regression logarithmic equation is given in Equation (2). The corresponding regression curve is shown in Fig. 5A.

The regression analysis results for the regular search were as follows: The  $R$  square was 0.897, the  $F$  value was 854.268, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.971$ ,  $b_1 = -0.107$ ). The regression curve is shown in Fig. 5B.

The regression analysis results for the exact phrase search are as follows: The  $R$  square was 0.874, the  $F$  value

was 677.898, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.860$ ,  $b_1 = -0.097$ ). The regression curve is shown in Fig. 5C.

The regression analysis results for the PDF file restriction search were as follows: The  $R$  square was 0.857, the  $F$  value was 585.852, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.973$ ,  $b_1 = -0.090$ ). The regression curve is shown Fig. 5D.

All the  $R$  square values of the four different types of search ranged from 0.845 to 0.897. The regular search

Table 8 Descriptive summary of the Live Search data.

Variable	N	Mean	Std. deviation	Std. error	95% Confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
Ti	20	0.6091	0.02371	0.00530	0.5980	0.6202	0.57	0.64
Re	20	0.5440	0.02691	0.00602	0.5314	0.5566	0.48	0.58
Ph	20	0.4561	0.03806	0.00851	0.4383	0.4739	0.34	0.51
PDF	20	0.6485	0.02905	0.00650	0.6349	0.6621	0.58	0.69
Total	80	0.5644	0.07894	0.00883	0.5469	0.5820	0.34	0.69

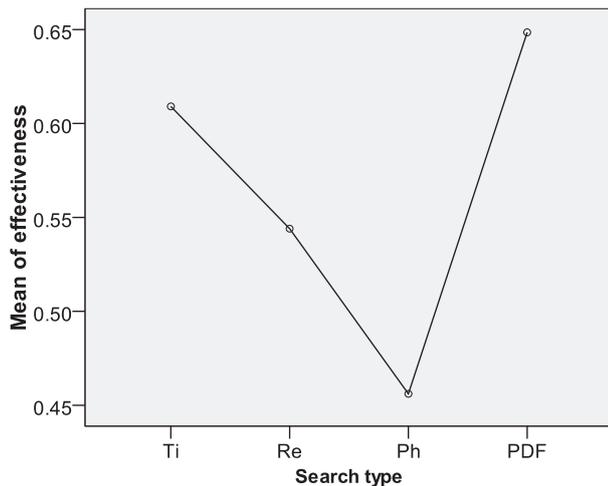


Figure 6 Mean plots for the Live Search test.

achieved the best regression curve fit (or web page ranking) due to its higher *R* square value (0.897). The web pages' distribution along the regression curve in Fig. 5B also supported this claim. The title search achieved the worst because of its lower *R* square (0.845).

It is interesting that the retrieval performance pattern for the four search features in *Yahoo* was quite similar to that for the four search features in *Google*. The *PDF* file format restriction search achieved the best retrieval performance while the exact phrase search achieved the worst retrieval performance in both *Yahoo* and *Google*. Their means plots charts looked similar. The regression analysis indicated that the regular search had the best regression curve fit in both *Yahoo* and *Google*.

### Data results and data analysis for Live Search

The one-way *ANOVA* test results for *Live Search* showed that *DF* was 3, the *F* value was 158.101, the mean square was 0.141, and the *p*-value/*sig* was 0.000 which is smaller than 0.05. Therefore, the null hypothesis *H3* was rejected. In other words, there were statistical differences among

regular searches, title searches, exact phrase searches, and *PDF* file format restriction searches in terms of search effectiveness in *Live Search*.

Table 8 shows the descriptive summary of the experimental study for *Live Search*. In the first column "Variable", the definitions of *Ti*, *Re*, *Ph*, and *PDF* are the same as Table 2. The *Live Search* test mean plots are presented in Fig. 6 where the definitions of the X-axis and Y-axis are the same as these in Fig. 4. The retrieval effectiveness mean of the exact phrase search (0.4561) was the smallest while that of the *PDF* file format restriction search (0.6485) was the greatest. The smallest standard deviation of the title search (0.02371) indicates that the regular search effectiveness was the most stable one compared to other search types. The greatest standard deviation was 0.03806 (the exact phrase search). Notice that the lower boundary and upper boundary for the *PDF* file format restriction search were 0.6349 and 0.6621 at 95% confidence interval, respectively. They were the highest scores in the lower boundary and upper boundary categories, respectively. The minimum and maximum means for the *PDF* file format restriction search were 0.58 and 0.69, respectively. They were also the highest scores in the minimum and maximum mean categories, respectively. On the other hand, notice that the lower boundary and upper boundary for the exact phrase search were 0.4383 and 0.4739 at 95% confidence interval, respectively. They were the lowest scores in the lower boundary and upper boundary categories, respectively. The minimum and maximum means for the exact phrase search were 0.34 and 0.51, respectively. They were also the lowest scores in the minimum and maximum mean categories.

The null hypothesis *H3* was rejected because of the low *p*-value (0.000). Therefore, a *Tukey HSD* method was employed to detect the reasons for the rejection. The resultant *Tukey HSD* analysis results are presented in Tables 9 and 10. The definitions of the factors (*I* and *J*) *Ti*, *Re*, *Ph*, and *PDF* in Table 9 and 10 are the same as these in Table 8. The mean difference (*I* – *J*) column in Table 9 shows that all the mean differences between the *PDF* file restriction search and other types of search were positive and significant: 0.03940 (the mean of *PDF* file format restriction

Table 9 Multiple comparisons in Tukey HSD for the Live Search test.

(I) Factor	(J) Factor	Mean Difference (I – J)	Std. error	Sig.	95% Confidence interval	
					Lower bound	Upper bound
Ti	Re	0.06510*	0.00946	0.000	0.0403	0.0899
	Ph	0.15300*	0.00946	0.000	0.1282	0.1778
	PDF	-0.03940*	0.00946	0.000	-0.0642	-0.0146
Re	Ti	-0.06510*	0.00946	0.000	-0.0899	-0.0403
	Ph	0.08790*	0.00946	0.000	0.0631	0.1127
	PDF	-0.10450*	0.00946	0.000	-0.1293	-0.0797
Ph	Ti	-0.15300*	0.00946	0.000	-0.1778	-0.1282
	Re	-0.08790*	0.00946	0.000	-0.1127	-0.0631
	PDF	-0.19240*	0.00946	0.000	-0.2172	-0.1676
PDF	Ti	0.03940*	0.00946	0.000	0.0146	0.0642
	Re	0.10450*	0.00946	0.000	0.0797	0.1293
	Ph	0.19240*	0.00946	0.000	0.1676	0.2172

\*The mean difference is significant at the 0.05 level.

**Table 10** Homogeneous subsets for the Live Search test.

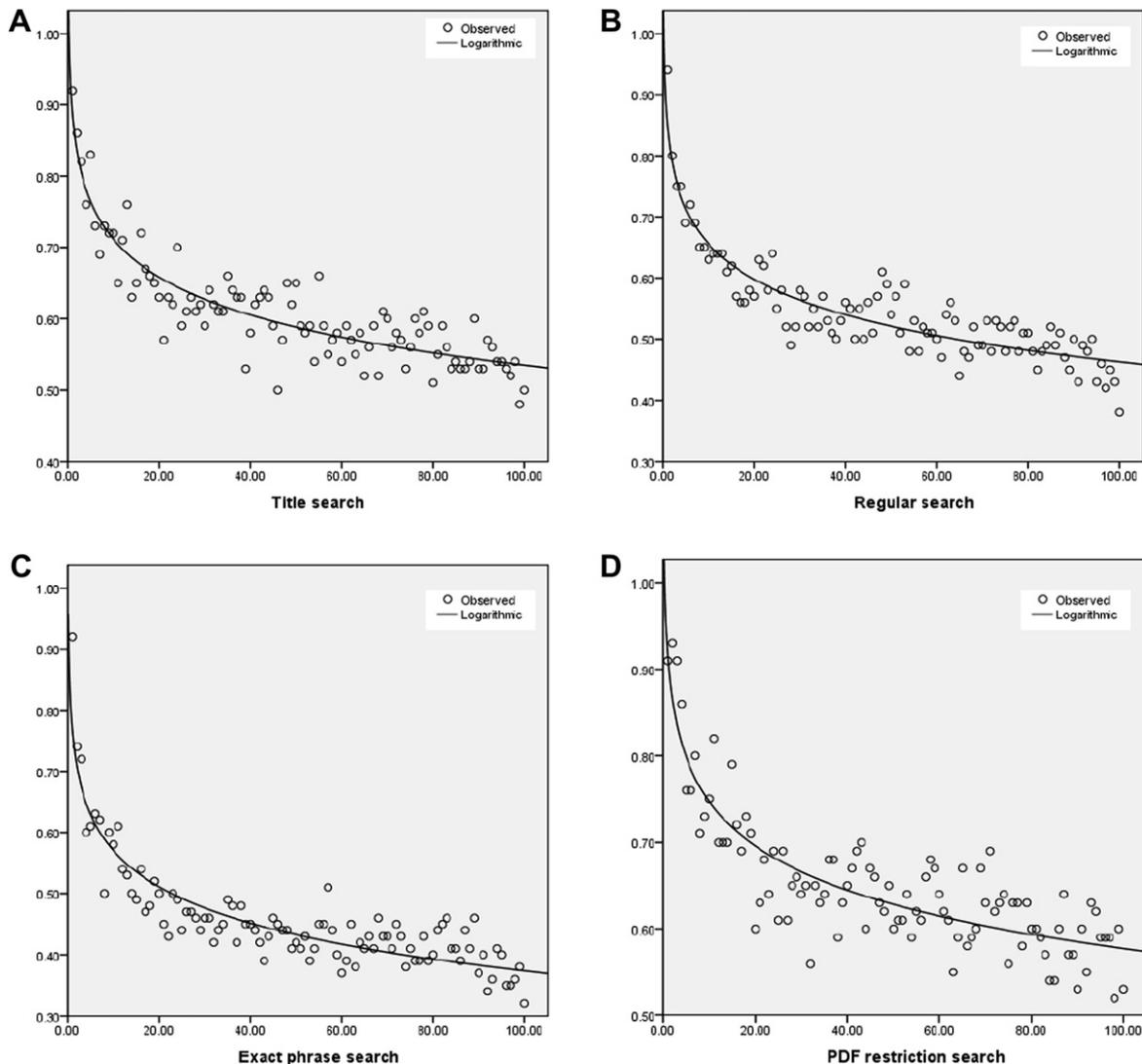
Factor	N	Subset for alpha = 0.05			
		1	2	3	4
Ph	20	0.4561			
Re	20		0.5440		
Ti	20			0.6091	
PDF	20				0.6485
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

search – the mean of title search), 0.10450 (the mean of *PDF* file format restriction search – the mean of regular search), and 0.19240 (the mean of *PDF* file format restriction search – the mean of exact phrase search). In other words, the *PDF* file format restriction search achieved the best retrieval performance among the four groups. The title search was in second place because the significant mean difference between the title search and the regular search was 0.06510, and the significant mean difference between

the title search and exact phrase search was 0.15300. The regular search was in third place because the significant mean difference between the regular search and the exact phrase search was 0.08790. It is evident that the performance of the exact phrase search was the worst among the four search groups. All the *sig* values for the mean differences are 0.000 in Table 9.

In the *Tukey HSD* analysis four separate homogeneous subsets were produced in Table 10. That is, each search



**Figure 7** Regression analysis for four search types in Live Search.

type corresponds to a homogeneous subset, which confirms that there were significant differences among the four search types.

A regression analysis was conducted for each of the four search types. In Fig. 7, the definitions of the X-axis and Y-axis are the same as those in Fig. 3.

The regression analysis results for the title search were as follows: The  $R$  square was 0.813, the  $F$  value was 426.066, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.888$ ,  $b_1 = -0.077$ ). The regression logarithmic equation is given in Equation (2). The corresponding regression curve is shown in Fig. 7A.

The regression analysis results for the regular search were as follows: The  $R$  square was 0.838, the  $F$  value was 507.456, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.849$ ,  $b_1 = -0.084$ ). The regression curve is shown in Fig. 7B.

The regression analysis results for the exact phrase search were as follows: The  $R$  square was 0.823, the  $F$  value was 454.702, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.766$ ,  $b_1 = -0.085$ ). The regression curve is shown in Fig. 7C.

The regression analysis results for the *PDF* file restriction search were as follows: The  $R$  square was 0.755,  $F$  value was 301.666, and  $Sig$  was 0.000. The regression equation was a logarithmic model ( $b_0 = 0.917$ ,  $b_1 = -0.074$ ). The regression curve is shown in Fig. 7D.

The  $R$  square values of all the four search types ranged from 0.755 to 0.838. The regular search achieved the best regression curve fit (or web page ranking) due to its higher  $R$  square value (0.838). The web pages' distribution along the regression curve in Fig. 7B also supports this claim. The *PDF* search achieved the worst performance in this category because of its lower  $R$  square (0.755). Fig. 7D also confirms this.

In summary, in all the three investigated search engines (*Yahoo*, *Google*, and *Live Search*) the retrieval effectiveness performance patterns for the four kinds of search shared commonalities. The *PDF* file format restriction search outperformed the other three search types in terms of retrieval precision, the title search was in second place, the regular search was in third place, and the exact phrase search achieved the worst retrieval performance. The regression analysis showed that the regular search had the best regression curve fit in all the three search engines, which means that result ranking for regular search was more accurate than the other three search types. However, the best regression curve fits in the title search, exact phrase search, and *PDF* restriction search did not behave consistently in the three search engines. The exact phrase search achieved two second places and one fourth place in terms of result ranking performance. The *PDF* restriction search achieved two third places and one fourth place. The title search achieved one second place, one third place, and one fourth place.

## Conclusion

Information seeking on the Internet is not only crucial for researchers, educators, engineers, and professionals but also vital for common people. Search engines are widely recognized as a primary means for information search on

the Internet. A search engine usually provides searchers with multiple search features/options to meet various users' information needs. Investigation on how search engines respond to their search features would assist users in developing appropriate search strategies and generating more satisfactory search results.

Toward this aim, three popular search engines (*Yahoo*, *Google*, and *Live Search*) were selected for the study. Advanced search features such as title search, exact phrase search, and *PDF* file format restriction search in these search engines were identified and their retrieval performance was analyzed in terms of retrieval effectiveness and retrieval result ranking performance. In addition, the regular search was chosen as a comparison baseline. In this experimental study, 240 searches were submitted to the three search engines (*Yahoo*, *Google*, and *Live Search*). As a result, 24,000 returned web pages were examined and their relevance judgments were made. Retrieval effectiveness for each of the 240 searches was calculated based on the defined precision.

The experimental results show that there were significant differences among the regular searches, title searches, exact phrase searches, and *PDF* file format restriction searches in terms of retrieval effectiveness in *Yahoo*, *Google*, and *Live Search*, respectively. Advanced search features such as the *PDF* file format restriction search and the title search outperformed the regular search in terms of retrieval effectiveness in the three search engines. It is surprising that the exact phrase search did not outperform the regular search in the investigated search engines. The advanced search features behavior patterns were similar across the three search engines.

The regression analysis results for the three search engines show that the regular search outperformed the title search, exact phrase search, and the *PDF* file format restriction search in terms of retrieval results ranking. The results are at odds with our expectations. In summary, the regular search outperformed the advanced search features in terms of retrieval result ranking while the advanced search features outperformed the regular search in terms of retrieval effectiveness. Some advanced search features (Title search and *PDF* file restriction format search) achieved better performance than the regular search. However, another advanced search feature, exact phrase search, did not outperform the regular search in either retrieval effectiveness or retrieval ranking performance.

The findings of this study suggest that:

- Keywords in the title of a Web page are more important and relevant to its contents, and using the title search feature in a search engine may improve the retrieval effectiveness;
- Documents in the *PDF* format are more formal than regular Web pages and they are usually official documentations, procedures, and publications. As a result using the *PDF* format search feature in a search engine would increase the opportunity that this kind of the documents are retrieved; and
- Due to the poor performance of the exact phrase search feature, search engine companies should revisit their algorithms for the exact phrase search feature by considering keywords in a phrase separately.

Selection of the examined features for a search engine was primarily affected by two factors: the nature of the search features and the availability of these features in the selected search engines. In this study, the three popular search engines were selected for comparison so that the behaviors of the search features could be identified among the three search engines. As a result, each of the selected features needed be available in the three search engines. Regular keyword search is widely used as the default search in search engines. In addition, the results of regular keyword searches served as a baseline when the results of another advanced feature were analyzed. A title is a general, distinctive, and descriptive heading of a web page. It represents the primary theme of a web page. Keywords appearing in a title should be more important than the keywords appearing other places in a web page. It is natural for users to select this title search feature. Exact phrase matching search does not treat the keywords separately in a query and the order of the keywords in a query is kept during the keyword matching process. Therefore, the possible ambiguity caused by different keyword combinations can be avoided. It is supposed to be used for a more specific search. PDF documents have many advantages over other document formats: They can be used across various platforms and they are independent of the hardware, operating systems, and applications; they offer document level security; and they also meet legal document requirements. With emergency of the new digitalization technology, more and more documents in the PDF format are available on the Internet. As a consequence, the PDF file format restriction search has become more important for users.

In returned results list from the search engines, there were few dead links. These dead links were excluded from the study. The rates of dead links in retrieval results lists for Yahoo, Google, and Live Search were 1.70%, 1.20%, and 2.70% respectively. Dead links in a retrieval results list can be caused by related server changes, directory path changes, or web page changes. During the relevance judgment process, it is difficult to judge the relevance of a web page based only on a web page title without reading its contents. That is why the dead links were not considered in the study. Because the dead link rate in the retrieval results list was very low, the impact of the dead links on the study results was not significant.

The search engine Live Search ceased in 2009 and was replaced by a new search engine, Bing. Because our experimental study on the three search engines was conducted from October 2008 to February 2009, Bing was not selected for this study.

Future directions on this topic include, but are not limited to, using the same research method to test more other advanced search features in these search engines, such as Boolean search, language restriction, other file type restriction, and domain restriction or to examine advanced search features in other search engines.

## References

- Bar-Ilan, J. (2006). Web links and search engine ranking: the case of Google and the query "Jew". *Journal of the American Society for Information Science and Technology*, 57(12), 1581–1589.
- Bar-Ilan, J., Keenoy, K., Yaari, E., & Levene, M. (2007). User rankings of search engine results. *Journal of the American Society for Information Science and Technology*, 58(9), 1254–1266.
- Crudge, S. E., & Johnson, F. C. (2004). Using the information seeker to elicit construct models for search engine evaluation. *Journal of the American Society for Information Science and Technology*, 55(9), 794–806.
- Eastman, C., & Jansen, B. J. (2003). Coverage, relevance, and ranking: the impact of query operators on Web search engine results. *ACM Transactions on Information Systems*, 21(4), 383–411.
- Fattahi, R., Wilson, C. S., & Cole, F. (2008). An alternative approach to natural language query expansion in search engines: text analysis of non-topical terms in Web documents. *Information Processing & Management*, 44(4), 1503–1516.
- Ford, N., Miller, D., & Moss, N. (2002). Web search strategies and retrieval effectiveness: an empirical study. *Journal of Documentation*, 58(1), 30–48.
- Hassan, I., & Zhang, J. (2001). Image search engine feature analysis. *Online Information Review*, 25(2), 103–114.
- Jacsó, P. (2008). Savvy searching: how many Web-wide search engines do we need? *Online Information Review*, 32(6), 860–865.
- Jansen, B. J., & Molina, P. R. (2006). The effectiveness of Web search engines for retrieving relevant ecommerce links. *Information Processing & Management*, 42(4), 1075–1098.
- Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management*, 42(1), 248–263.
- Kumar, R., Suri, P. K., & Chauhan, R. K. (2005). Search engines evaluation. *DESIDOC Bulletin of Information Technology*, 25(2), 3–10.
- Leighton, H. V., & Srivastava, J. (1999). First 20 precision among World Wide Web search services (search engines). *Journal of the American Society for Information Science*, 50(10), 870–881.
- Mowshowiz, A., & Kawaguchi, A. (2005). Measuring search engines bias. *Information Processing and Management*, 41(5), 1193–1205.
- Nicholson, S., Sierra, T., Eseryel, U. Y., Park, J. H., Barkow, P., Pozo, E. J., et al. (2006). How much of it is real? Analysis of paid placement in Web search engine results. *Journal of the American Society for Information Science and Technology*, 57(4), 448–461.
- Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2003). Multimedia Web searching trends: 1997–2001. *Information Processing & Management*, 39(4), 611–621.
- Pew Research Center. (2008). Internet activities. Retrieved June 11, 2009, from <http://www.pewinternet.org/Data-Tools/Download-Data/~ /media/Infographics/Trend%20Data/January%202009%20updates/Internet%20Activities%20-%20all%20-%201%206%2009.jpg>.
- Spink, A. (2002). A user-centered approach to evaluating human interaction with Web search engines: an exploratory study. *Information Processing & Management*, 38(3), 401–426.
- Spink, A., Jansen, B. J., Blakely, C., & Koshman, S. (2006). A study of results overlap and uniqueness among major Web search engines. *Information Processing & Management*, 42(5), 1379–1391.
- Su, L. T. (2003). A comprehensive and systematic model of user evaluation of Web search engines: I. Theory and background. *Journal of the American Society for Information Science and Technology*, 54(13), 1175–1192.
- Sullivan, D. (2006). ComScore Media Metrix search engine ratings. *Search Engine Watch*. Retrieved June 11, 2009, from <http://searchenginewatch.com/2156431>.

- Thelwall, M. (2008). Quantitative comparisons of search engine results. *Journal of the American Society for Information Science and Technology*, 59(11), 1702–1710.
- Vaughan, L., & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 40(4), 693–707.
- Wallop, H. (2009). More than 1 billion internet users now online. *The Telegraph*. Retrieved May 7, 2009, from <http://www.telegraph.co.uk/scienceandtechnology/technology/4325987/More-than-1-billion-internet-users-now-online.html>.
- Wu, L. L., & Chuang, Y. L. (2008). Motivation for using search engines: a two-factor model. *Journal of the American Society for Information Science and Technology*, 59(11), 1829–1840.
- Zhang, J., & Chueng, C. (2003). Meta search engine feature analysis. *Online Information Review*, 27(6), 433–441.
- Zhang, J., & Dimitroff, A. (2005). The impact of webpage content characteristics on webpage visibility in search engine results (part 1). *Information Processing & Management*, 41(3), 665–690.
- Zhang, J., & Lin, S. (2007). Multiple language supports in search engines. *Online Information Review*, 31(4), 516–532.