International Journal of Information Science and Management
Vol. 17, No. 1, 2019, 47-57

Search Strategies and the Relevance of Retrieved Information in Persian Articles Database: Survey of M.A Students of Shiraz University

Hassan Moghaddaszadeh
Assistant Prof. in Knowledge and Information Sciences
Secretary of Iranian Library and Information Science, Association-Fars Branch
moghaddas1354@gmail.com

Abstract

Retrieving relevant information on the Internet and identifying the related information to the real needs are not an easy task for many users. So the main objective of this study was to evaluate the effect of search strategies on the relevance of retrieved information in domestic article databases. Considering the nature of the subject, this was an applied descriptive-survey research. Statistical population consists of all domestic article databases, from which the MAGIRAN, IRANDOC, NOORMAGZ and the Regional Information Center for Science and Technology (RICeST) were selected as samples. To test the hypotheses, one-way analysis of variance (ANOVA) and Tukey’s post-hoc test were computed using SPSS statistical software version 22. The study’s findings showed that there were significant differences between relevance of the information retrieved from different databases based on different search strategies. It was found that, using simple search had the highest relevance. Moreover, using the AND, NOT and OR operators, took the lower ranks respectively. Using the time limiter had the lowest relevance in information retrieval. There were also significant differences between the relevance of information retrieved from different databases, and the NOORMAGZ database, the RICeST, MAGIRAN and IRANDOC; respectively had the most relevant retrievals. Using different search strategies can affect the relevance of the information retrieved from an article database. Therefore, acquiring these strategies and using each one in the right situation can improve the relevance of the retrieved information.

Keywords: Information Retrival, Relevance, Search Strategy, Article Database, Magiran, IRANDOC, Noormagz, RICeST**.**

Introduction and Problem Statement

The world of Web huge volume of information. In order to take advantage of the information available on the Web, the user should search through efficient and suitable methods. This task (i.e. information retrieval) is a complicated one since it depends on different aspects of searching the databases and search engines. Moreover, during the retrieval of data from the Web, understanding the proper methods of searching and their steps are highly significant. This could be realized via proper systems and models of searching for and retrieval of data (Snasel, Abraham, Owais, Platos, & Kromer, 2009). Proper use of searching systems and strategies can contribute to retrieve relevant information by user in response to his/her requirement.

On the World Wide Web, retrieval of information is done via searching tools (i.e. search engines and thematic guides) as well as specialized websites and databases. Through indexing web pages, search tools inform the researcher of existence of data in different locations. Searching tools usually offer a list of tens of thousand of pages in response to a search request (i.e., queries). Each searching tool has its own unique content, coverage and interface. For using them and conducting a successful and useful search, one should be aware of a set of searching principles and rules that those tools adopt. The principles are called information search strategies. Information search strategy is a method adopted by the searcher for finding information on the web space. If a researcher does not know the system, solutions and ways of accessing online information, he/she will not be able to obtain precise and relevant information in due time and efficient manner. Consequently, he/she will spend hours searching for information and ends up with piles of uncategorized information. As a result, every person (especially researchers) should be familiar with searching methods and principles so as to minimize the time of searching, and access information more methodically and use the information in their field of expertise (Habibi, 2007). A significant issue to be raised in terms of retrieval of data from the Internet is the relevance of retrieved data.

In simple terms, relevance is the degree of association between results retrieved from storage and retrieval system and the user's question. Data retrieval and storage systems seek to offer relevant data to the users. Because searching for data is aimed at decision making, if information is offered by data retrieval and storage system, system user receives it, and received information causes changes in an individual's knowledge structure, then one could state that offered information will be relevant for the recipient, data transfer will be effective and desirable results will be retrieved (Hassanzadeh and Reza-Zadeh, 2008).

In fact, the concept of relevance dates back to centuries ago and activities of earliest libraries. At the time, users of libraries sought to find relevant information too. As volume and significance of information have increased and automatic retrieval systems developed, relevance was achieved in more serious ways. As a result, different studies concerning relevance have been conducted. Ellis argued the division of these studies into cognitive, behavioral and affective aspects (as cited in Okhovati, 2004, P. 23). In fact, early studies were mostly cognitive but as time passed the studies turned to more user-oriented studies, and this field attracted more attentions. In the following, a summary of relevant studies is offered.

In order to review the previous studies, Fattahi (2006) conducted a survey on the identification and analysis of general terms in web resources based on a new approach so as to expand search term in search engines through natural languages. The results suggested that query expantion could be highly successful in the case of searching for term and online address (URLs) for each area of subject. As a result, search engines could limit default research to term-based and online address research.

In reviewing researches on the subject of this study and the relevance issue the results of

Shahbazi and Shahini (2016) entitled “study of the efficiency of Magiran, Noormags and SID databases in retrieval and relevance of information science and knowledge subject by free keywords and comparing them in terms of the use of controlled keywords” showed that there is a great difference regarding free and controlled keywords retrieval of information and knowledge science in Noormagz database compared with other two databases. In addition, studying the thematic relevance of research data showed that the ability of this database for receiving othe related articles is more than two other databases.

In addition, the online addresses (URLs) and their use in marketing could influence efficient information marketing results. Hassanzadeh, Ghafari, Zarei and Kamandi (2014) carried out a comparative investigation on searching through terms and online address in search engines with results of relevance of information retrieval from the viewpoint of experts of Hamedan Governorship. The results suggested that, as the number of search key termst increased, consistency of terms and online addresses reduced. In addition, the higher consistency between the search term and online address is correlated with higher relevance.

Allami and Fattahi (2012) conducted a study under the title “Comparison of the effect of term and online address on the relevance of results of data retrieval in search engines in two disciplines, agricultural sciences and humanities”, and concluded that the relevance of online address results is higher than term-based search in agriculture sciences. In humanities, there was no significant difference between relevance of search results based on term and online address. They concluded that users could attain more relevant results by controlling the number of search key terms. In addition, specialized search engines could rate more relevant results and offer them to users through giving more weight to the existing key terms in the online address.

Adoption of search strategies is another factor affecting efficient data retrieval. Badgett, Dylla, Megison and Harmon (2015) developed an experimental model for efficient searching of information in medical fields offered in Pubmed data base and Google search engine. The results suggested that when compared with other strategies, their experimental strategymodel has the highest output of relevant data retrieval. This strategy retrieved papers with highest quality and reference frequencies, so that comparison of all retrieval results through this model and other models pointed to a significant level of difference. On the other hand, the setting in which a person is engaged could be a significant factor in his/her judgment of relevance. This subject was studied by Barral et al. (2015). The results of their study suggested that facilities, form and structure of the page retrieved in 4 to 6 seconds could influence users’ relevance judgment. Therefore, the webpage space could effectively affect users’ judgment about relevance, and webpage designers should take this problem into attention.

A review of structured search, linguistic models and relevance models in data retrieval languages was carried out by Larkey and Connell (2005). To do this, they adopted two methods to search for Arabic and Spanish words in dual-language dictionaries of common words and derived words. The results suggested that structured searches offers better and more relevant results. On the other hand, when demands extend linguistic models offer better results.

The literature review suggested that searching based on online addresses (URLs) could offer higher relevance than termbased searching. Most people looked for information by simple keyword search. They use thematic guide and advanced search less frequently. The findings suggested that if searchers have exact online address for accessing a Webpage, paper or data, they could access their information on the Web in an easier and more precise manner. This could improve the relevance of data retrieval. On the other hand, advanced search options in search engines could offer more relevant results due to application of different limitations (i.e. time. language, etc.). However, use of these options requires searching skills. On the other hand, thematic guides in search engines (e.g. Yahoo) could systematically guide an individual toward desirable results if the person is aware of subject`s divisions in each field.

Based on the results of previously conducted studies and the increasing volume of data on the World Wide Web, the retrieval of relevant data out of piles of existing information will be increasingly significant. The significant question to be raised here is which and to what extent different strategies of searching could affect the relevance of retrieved data? In addition, is there a difference between frequently used Persian articles database in terms of relevance of retrieved information?

The problem was addressed in the present survey from viewpoints of students of six higher education majors who were studying in departments of education sciences and psychology at Shiraz University, Iran. The results and findings of the present study could help designers and managers of Persian databases plan their retrieval systems optimally. In addition, users of these databases could draw upon these results to choose the desirable search strategy for each database. In this case, they can improve the relevance of retrieved information and obtain more desirable and useful information.

The review of the related literature and theoretical principles led to the following hypotheses for the present survey.

Main Hypothesis

There is a significant difference between the relevance of the information retrieved from the Persian articles database with regard to different search strategies.

SecondaryHypothesis

1. There is a significant difference between the relevance of retrieved information from different databases.

2. There is a significant difference between the relevance of retrieved information using Boolean operators in databases.

3. There is a significant difference between the relevance of the information retrieved using the time limiter in databases.

Methodology

Considering the nature of the research subject, the present study is a descriptive survey and one may conclude that it is applied in term of its objective. The statistical population of present study includes all Persian articles databases which, due to the impossibility of reviewing all of them, more relevant and more inclusive ones were selected as sample: MAGIRAN, IRANDOC, NOORMAGZ and the Regional Information Center for Science and Technology (RICeST). The second statistical population included M.A. students of Education and psychology in Shiraz University. This field of study was selected due to nature of surveys, diversity of subjects and coverage of their subjects of study by all reviewed databases. Due to the nature of the present survey, a small but representative group of users (30 MA students) were selected out of the whole statistical population so as to determine the level of relevance of results obtained through the chosen databases. The group included those university students who were writing proposals or MA thesis at the time and who were volunteered to cooperate with researcher. The reason behind limiting statistical population to this group of students was their high need for relevant information and resources to complete writing their thesis, and they could cooperate with the researcher and search for information in a real space.

In the present survey, the pooling method was adopted as described in the following. First, search terms were given to a number of searchers. After searching in data retrieval system, each searcher should offer an inventory of retrieval results including 20 first retrieved records for each search term. Then, inventories of search items were integrated into a repository. The records not included in the repository were considered to be irrelevant. Based on their information requirements, the selected students searched in intended databases. Then, they searched for five key terms through natural language, Boolean operators and time limiters. Therefore, the only search strategy used in the present survey was using search box in the advanced search section of the intended database which had been limited by Boolean operators and time limiters. After searching, the participants evaluated the relevance of retrieved items (first 20 results) with information requirement as very low (1), low (2), medium (3), high (4) as well as very high (5). For data analysis, descriptive statistical measures of mean and standard deviation were used. In the section of inferential statistics, one-way ANOVA and Tukey follow-up tests were conducted for comparison of relevance of retrieved data based on search strategies and comparison of databases. To do these tasks, SPSS Software (version.22) was used.

Data Analysis

In order to determine the type of test used to prove or reject the hypotheses for the study, first, we have examined the normal or abnormal distribution of data, based on the results obtained from a suitable statistical test from parametric and nonparametric statistical tests is used to test the research hypotheses.

Table 1

Test the status of the data distribution

Search Strategi	P-Value	α	Result
Simple	0.061	0.05	Normal
AND	0.161	0.05	Normal
OR	0.055	0.05	Normal
NOT	0.494	0.05	Normal
Time limitation	0.057	0.05	Normal

As shown in table 1, in all variables, the significance level (P) is greater than the error level (0.05), which indicates the normal distribution of data, therefore one-way analysis of variance (ANOVA) and Tukey's post hoc test were used to test the research hypotheses.

Main Hypothesis

There is a significant difference between the relevance of the information retrieved from the Persian articles database with regard to different search strategies.

Table 2

Compare the relevance of retrieved information from Persian articles database according to the search strategy

	Simple		AND		OR		NOT		Time Limitation		Test	
	M	SD	M	SD	M	SD	M	SD	M	SD	F	P
Rel	65.70	14.64	62.85	10.67	46.65	7.43	43.50	5.24	17.70	8.45	59.020	0.0001

Based on Table 2 and level of significance (p=0.0001) which is less than critical value (i.e. 0.05), zero hypothesis was denied and research hypothesis supported. Therefore, with 95% level of confidence, one could state that there was a significant difference between mean relevance of data retrieved from intended databases according to different search strategies, so that a simple search offered the highest level of relevance followed by use of AND operator, OR operator, NOT operator and Time limiter offered the least relevance of data retrieval. Considering the significant difference between different search strategies in terms of relevance of retrieved data, Tukey follow-up test was used to compare each pair of search strategies. The results are shown in Table 3.

Table 3

Tukey test results in comparison between search strategies

Group	Comparison Group	Mean Difference	SD Error	P
Simple	AND	2.85	3.26	0.958
	OR	19.05	3.26	0.0001
	NOT	22.20	3.26	0.0001
	Time Limitation	48.00	3.26	

Group	Comparison Group	Mean Difference	SD Error	P
AND	Simple	-2.85	3.26	0.958
	OR	16.20	3.26	0.0001
	NOT	19.35	3.26	0.0001
	Time Limitation	45.15	3.26	0.0001
OR	Simple	-19.05	3.26	0.0001
	AND	-16.20	3.26	0.0001
	NOT	3.15	3.26	0.936
	Time Limitation	28.95	3.26	0.0001
NOT	Simple	-22.20	3.26	0.0001
	AND	-19.35	3.26	0.0001
	OR	-3.15	3.26	0.936
	Time Limitation	25.80	3.26	0.0001
Time Limitation	Simple	-48.00	3.26	0.0001
	AND	-45.15	3.26	0.0001
	OR	-28.95	3.26	0.0001
	NOT	-25.80	3.26	0.0001

As data of Table 3 suggests, there was no significant difference between the relevance of data retrieved from simple search and use of AND operator. In addition, no significant difference was observed between use of OR and NOT operators and relevance of retrieved information. In other cases, pairwise comparison of search strategies suggested a significant difference between all of them.

H_1: There is a significant difference between the relevance of retrieved information from different databases.

Table 4
Compare the relevance of recovered information from databases

Database	IranDoc		Magiran		Noormags		RICeST		Test	
	M	SD	M	SD	M	SD	M	SD	F	p
Relevance	8.59	4.33	9.91	6.86	17.22	5.65	14.15	8.72	43.249	0.0001

Acording to Table 4 and level of significance (p=0.0001) which is less than critical value (i.e. 0.05), zero hypothesis was denied and research hypothesis supported. Therefore, with 95% level of confidence, one could state that there was a significant difference between mean relevance of data retrieved from intended databases, so that Noormags search database offered the highest level of relevance followed by RICeST database, Magiran database and IranDoc database.Considering the significant difference between different databases in terms of relevance of retrieved data, Tukey follow-up test was used to compare each pair of databases. The results are shown in Table 5.

Table 5

Tukey test results in comparison between databases

Group	Comparison Group	Mean Difference	SD Error	P
IranDoc	Magiran	-1.32	0.85	0.411
	Noormags	-8.63	0.85	0.0001
	RICeST	-5.56	0.85	0.0001
Magiran	IranDoc	1.32	0.85	0.411
	Noormags	-7.32	0.85	0.0001
	RICeST	-4.24	0.85	0.0001
Noormags	IranDoc	8.63	0.85	0.0001
	Magiran	7.32	0.85	0.0001
	RICeST	3.07	0.85	0.002
RICeST	IranDoc	5.56	0.85	0.0001
	Magiran	4.24	0.85	0.0001
	Noormags	-3.07	0.85	0.002

As showed on Table 5, there was no significant difference between the relevance of data retrieved from the IranDoc and Magiran. In other cases, the comparison between databases in two to two indicates a significant difference between the relevance of the information retrieved from them.

H_2: There is a significant difference between the relevance of retrieved information from databases using Boolean operators

Table 6

Compare the relevance of recovered information form databases using Boolean operators

Engine / Bolian	IranDoc		RICeST		Magiran		Noormags		Test	
	M	SD	M	SD	M	SD	M	SD	P	F
AND	8.10	4.54	22.30	2.83	12.60	4.59	19.85	2.32	61.995	0.0001
OR	7.70	3.18	8.65	3.50	10.10	2.31	20.20	3.43	67.650	0.0001
NOT	13.00	3.46	11.10	1.74	7.85	2.24	11.55	1.64	16.272	0.0001

Based on Table 6 and level of significance (p=0.0001) which is less than critical value (i.e. 0.05), zero hypothesis was denied and research hypothesis supported. Therefore, with 95% level of confidence, one could state that there was a significant difference between mean relevance of data retrieved from intended databases according to different Boolean operators, so that AND operator had most relevance recovered information at RICeST database, followed by Noormags, Magiran and IranDoc database. Data showed that with use of OR operator Noormags retrived best relevance information and it followed by Magiran, RICeST and IranDoc. With use of the NOT operator for search information IranDoc retrived most relevant information, after that were Noormags, RICeST and Magiran.

H$_3$. There is a significant difference between the relevance of the information retrieved from the databases using the time limiter.

Table 7

Compare the relevance of recovered information from databases using time limiter

Database	IranDoc		RICeST		Magiran		Noormags		Test	
	M	SD	M	SD	M	SD	M	SD	P	F
Relevance	6.05	2.11	---	---	1.30	1.38	10.35	6.11	40.872	0.0001

Acording to Table 7 and level of significance (p=0.0001) which is less than critical value (i.e. 0.05), zero hypothesis was denied and research hypothesis supported. Therefore, with 95% level of confidence, one could state that there was a significant difference between mean relevance of data retrieved from intended databases with using time limiter. Only RICeST database hadn’t any time limitation in its search box, so to compare with other search engines didn’t earned any mean score. Among the three other databases, Normgs with a mean score of 10.35 had the best situation was followed by IranDoc respectively with a mean of 6.05 and Magiran with an average of 1.30.

Discussion and Conclusion

The findings of the present survey suggested that there was a significant difference between the relevance of data retrieved from articles database according to different search strategies. The use of simple search offered the highest relevance level of retrieved data followed by the AND operator. The time limiter achieved the least relevance of data retrieval. As Habibi (2007) suggested, simple and keyword searches were more frequently used by users. In addition, Fattahi (2006) found out that use of search strategies and query expantion in Google search engine could result in more precise and relevant information. The results and findings of present study suggest that use of different search strategies could result in different results. These strategies and proper use of them could direct users towards more relevant results. The adoption of limiters and limiting strategies (e.g. some Boolean operators) could reduce the amount of results but offer more relevant documents for satisfying user’s requirements. Because the present survey reviewed just the top 20 documents, the use of time limiters to limit the search to documents published within the past 1 year led to the retrieval of less than 20 documents in some cases. On the other hand, just Google search engine had time limitation, and this issue influenced and reduced the total score of other four engines. Therefore, one could conclude that proper use of strategies and limiters could reduce the volume of retrieved data and help the user attain more relevant data.

The comparison between the relevance of the information retrieved from the article databases of the reviewed showed that the Noormagas database had the best status, followed by the RICeST, Magiran and, finally, IranDoc. The findings of the research by Shahbazi and Shahini (2015) indicated that the Noormags articles database, in comparison with Magiran and the SID, has provided more relevant information in the field of library and information science. In the interpretation of the findings of the present study, it can be stated that the

Noormags articles database has been designed and planned for articles only, and in the humanities, and because of its full and powerful search engine, it has managed to overcome it from other databases.

The RICeST database was ranked second in this ranking, due to the lack of time limit and the score of this variable has been able to affect the average score of this database.

In addition, the RICeST covers a variety of sources, which in this case only the article database has been reviewed. On the other hand, this database coveres all disciplines compare to Noormags. Also, the IranDoc database is designed to cover theses, although it also covers articles. Only the Magiran database of the is specially tailored to the articles of the journals, and it is interesting to have the lowest rating, due to its newness and the small number of records available in this database.

The findings of the present study and theoretical principles of using search strategies and different limiters of data search and retrieval suggested that adopting these techniques and strategies could affect relevance of retrieved data. Reduced or increased relevance is probably dependent on the user; his/her search skills and how and where he/she uses the limiters. Obviously, one of the most significant factors affecting data retrieval is user’s search skills and proper use of techniques of searching in different engines and databases. In addition, familiarity with different databases and subject fields and resources the databases use could affect the relevance of information retrieved by search engines. Therefore, educating the information literacy skills in general and educating the search skills in particular could affect the relevance of retrieved data. On the other hand, designers of search engines could create new options and search facilities in the advanced search section of their products and help users retrieve less but more relevant data. This adds to user’s willingness to use these engines and encourages others to use the same search tools.

Considering the findings of this research and the considerations that the researcher encountered during the research, the following suggestions are presented:

1. It is suggested to the RICeST database administrators add a time limit in the advanced search section of their database and in addition to include the NOT function as the default in this section.
2. The accurately indexing of articles in databases to retrieve specific information related to keyword choices and attention to conversational language at this stage .
3. Use of efficient software in the field of searching and retrieving information indexed in the database.
4. Visibility of the database through public search engines such as Google, Yahoo and so on.
5. Benefiting frome information and knowledge science professionals in indexing and storing information.
6. Use of information and science experts as advisor in software designing for searching and retrieving information by databases.

References

Allami, P. & Fattahi, R. (2012). Comparing the Influence of Title and URL in Information Retrieval Relevance in Search Engines Results between Human Science and Agriculture Science. *Iranian Journal of Information Processing & Management*, 28 (1), 203-224. [in Persian]

Badgett, R. G., Dylla, D. P., Megison, S. D. & Harmon, E. G. (2015). An experimental search strategy retrieves more precise results than PubMed and Google for questions about medical interventions. *Peer J,* 3 (91), 1-15.

Barral, O. et al. (2015). Exploring peripheral physiology as a predictor of perceived relevance in information retrieval. *Proceedings of the 20th International Conference on Intelligent User Interfaces*, March 29- April 1, Atlanta. 389-399.

Fattahi, R. (2006). Identifying and analyzing general terms in web resources: a new approach to extending search terms unisng natural language in exploration engies. *Studies in Education and Psychology*, 7 (1), 31-51. [in Persian]

Habibi, S.H. (2007). The status of using information retrieval tools and search strategies for Ardabil University of medical sciences on the web. *Quarterly Book,* 69, 205-216. [in Persian]

Hassanzadeh, M. & Rezazadeh, E. (2008). Assessment of relevance in systems for storing and retrieving information from a cognitive approach. *Library and Information Science,* 11 (2), 53-70. [in Persian]

Hassanzadeh, M., Ghafari, S., Zarei, A. & Kamandi, H. (2014). Comparative review of search through the title and URL of search engines on the relevance of the results of data retrival by the experts of the hamedan governorate. *Quarterly Journal of Knowledge Studies,* 7 (25), 71-79. [in Persian]

Larkey, L. S. & Connell, M. E. (2005). Structured queries, language modeling and relevance modeling in cross- language information retrieval. *Information Processing and Management: An International Journal,* 41 (3), 457-473.

Okhovati, M. (2004). Concept of relevance in information retrieval systems: An overview of existing theories and literature. *Informology,* 2 (1), 23-45. [in Persian]

Shahbazi, M. & Shahini, S. (2016). Study of efficiency of magiran, noormags and sid databases in retrieval and relevance of information science and knowledge subject by free keywords and comparing them in terms of the use of controlled keywords. *Iranian Journal of Information Processing & Management,* 31 (2), 431-454. [in Persian]

Snasel, V., Abraham, A., Owais, S., Platos, J. & Kromer, P. (2009). Optimizing information retrieval using evolutionary algorithms and fuzzy inference system. *Foundations of Computational Intelligence,* 4 (204), 299-324.