



---

## Improving Ranking Persian Subjects in Search Engine Using Fuzzy Inference System

Elaheh Golzardi<sup>1\*</sup>, Majid Meghdadi<sup>2</sup> and Abdolbaghi Ghaderzade<sup>1</sup>

<sup>1</sup>Department of Computer Sciences, Research Branch, Islamic Azad University, Kurdistan, Iran

<sup>2</sup>Department of Computer Engineering, University of Zanjan, Iran

\*Corresponding Author's E-mail: [E.gol2412@gmail.com](mailto:E.gol2412@gmail.com)

### Abstract

According to the research, the efficiency of the search engines which done the rankings of Farsi content was much lower than the English search engines. After reviewing the literature, we found that, so far there been no ratings Persian system with fuzzy system and however, due to its proven performance in the field of fuzzy systems, also a search engine designed to accomplish this goal. Therefore, we prefer to advance this goal, so we establish a fuzzy inference system. It is created with the best evidence that can be considered to have been largely bringing the intended page to a user. Proposed method, display the relevant pages to the user in order to allow users to reach to their intended pages with less time and less cost. Also, in order to evaluate this method, Comparisons with other search engines was done.

**Keywords:** search engine, optimization, and ranking.

### 1. Introduction

Mostly general search engines provided accurate quantitative results, even for the exact query. Thus, extraction of useful information is a vital necessity. Today, search engines are important tools in the internet world for data types such as text, videos and photographs.



-----

While the results of the irrelevant information leads to wasted time and the cost for the user. This is while that the same user prefers to see the maximum two pages that displayed by the motor engine, because the results are likely to reach to her/his goal[1]. Because most researchers focus on the ranking content in English or Arabic, here our focus is on improving better rankings Persian content's system.

In this paper in order to improving the ranking we used a fuzzy inference system, which could help us to achieve the desired results in this field; Because a fuzzy inference system by using logics and rules can help us to reach our goal. The system is compared to other search engine which is able to response to the user in lower time and cost to deliver the best results in your search.

Considering that we use fuzzy logic to achieve our goal, so we need to know a short introduction of fuzzy logic. Theory of Fuzzy sets describe the uncertainty and imprecision in the event that the key idea of fuzzy logic is multi value logic. Fuzzy theory is supported Non-probabilistic uncertainty. Although theory of fuzzy sets are describes encoding, and a binary numbers. In this theory, join of members is shown with  $u(x)$  function that  $X$  represents a specific member and  $U$  is a fuzzy function that the degrees of membership  $X$  in fuzzy sets corresponding set of values between zero and one.

$$\tilde{A} = \{(x, \mu_A(x)) \mid x \in X\} \quad (1)$$

In other words,  $U(x)$  mapping of  $X$  may make numerical values between zero and one. Function may be discrete or continuous set of values; it is possible that  $U$  only some form of discrete values between zero and one, for example, including the numbers between zero and one are like 0.3 and 0.5. Or is continuo us, in this case a continuous curve represents decimal numbers between zero and one. Due to all these features and characteristics that exist in

-----

fuzzy logic, we would prefer to achieve top rankings so we establish a fuzzy inference system. A fuzzy system is knowledge based rule and its heart is a knowledge base of if - then rules. In this part a sample code that used in this article is taken.

*If (InputLinks is Bad) & (AvgMeets is Good)*

*& (Domain is VGood) then (Rank is VGood)* (2)

The design of fuzzy inference systems are used in various organ functions, the proposed method of Triangular Membership function was used; this is simply due to the calculation of the output membership functions most widely used in practice. Example of Join membership is shown in picture number 1.



Figure 1: Example of membership functions



-----

Because of the importance of ranking functions, as the main component of a search engine, much work has been done in this area with several techniques. In continue we have a Review on the Literature. In an article effective algorithms for improving performance of engine results have introduced. The two statistical methods based on ratios (Z-tests) and chi-square test (T-tests) was developed to extract the content of external. Also comparative studies between these two methods have been presented. Experimental results have shown better results than the statistical approach is used to generate chi-square test[2].

A paper check search engine ranking algorithms and variables are presented. This article is about the recognition of web-spam, hyperlinks analysis, and discussed the basic structure and duplicate content issues. In addition, the variables on the page and off-page search engine rankings in two categories of factors and their possible implications for ranking web documents are analyzed[3]. A comparative study was also provided by Google and Yahoo. The results of this study showed that the accuracy of simple multi-word queries for Google was high and also Yahoo has a relatively high precision for complex multi-word queries. Relative recall of a simple word in queries for Google was high, while Yahoo has a higher relative recall of complex multi-word queries[4].

An Arabic search engine ranking of module was developed that in this study, the authors focused on the implementation of an improved ranking algorithm. The combination of content and links cloud also is the root of this Arabic word in the text. Also the external database that contains the meanings of the Arabic words used most morphologically . Search results largely taken the form of a words that is belongs to that mean. This helps to reduce the inset[5]. In another article the difference between a Web ranking algorithms with statistical methods and tools presented. The aim of the present study confirms the idea of using a statistical method. To do this, five super search engine with four common search engines are used for this study is the use of five questions. These findings confirm that



-----

different search tools on the web are used to rank the different algorithms[6]. Also a stemming algorithm for Persian was wrote that authors in their paper bring the results of the design and implementation of a root finder for Persian language. In fact, they were the first overview of Farsi morphology and root-finding algorithm and its implementation that were evaluated to reach a conclusion on the right[7]. In another article, a relationship based on fuzzy clustering was proposed to explore user profiles for personalized web directory. User profiles can be used in a sequence of pages that the user's access increases, so the process can be used personalization. With personalization, web access or web page content changes. The proposed algorithm can give people information on the Web page for the visitor's personal website presentation[8].

A method based on fuzzy logic is presented. In this paper, a national quality assessment and performance measurement of dynamic Web sites is presented. That name is Fuzz-Web a natural and holistic way of reasoning based on multiple criteria decision-making process of the show. Then They try to use fuzzy logic as an intelligent technology, subjectivity and imprecision inherent volatility has been determined by the assessment process. Obviously, the selection of an appropriate evaluation criteria for the decision making process is necessary[9].

A fuzzy ontology-based search engine, fuzzy concept was created. In this paper, they have developed a search engine named Fuzzy-Go. First, a fuzzy ontology terms using fuzzy ontology to capture is made, then the appropriate semantic distance between semantic keyword search terms to perform is offers. Secondly, the users can insert multiple keywords with varying degrees of importance according to your needs. The third, provides category a range of web pages to users, who are prohibited from web pages in the areas, is poor. This reduces the search space and improves the search results[10]. In [11,12] Action to improve the ranking of each of our proposed algorithm and the results are expressed. Also in [13-16]



-----

the authors, by using different methods, have attempted to search engine optimization that can be used to improve better search engine.

## **2. Suggested work**

### **2.1 Software used**

In this research, to improve better rankings of Persian content in the search engines, our fuzzy systems is used in MATLAB. This ranking which is based on fuzzy theory is planned; because fuzzy logic has been help many researchers in many cases to reach their goals. In fact, it is clear that this logic; is a logic that don't see the world as a Binary numbers but sees as the gray spectrum. Therefore, we prefer to help with this issue would be to phase in this case to achieve better results and more accurate. The proposed system makes it possible for the user in the search space, to insert the words. Then based on the analysis of user requirement can be addressed by using the criteria defined in the system, the database search and selection will be final. After performing, a series of content's as close to the desired output is provided.

#### **2.1.1 Defining Input and Output:**

The input to the fuzzy system is planned, including:

- 1) The number of inbound links of each page that has links from other websites.
- 2) The number of daily users in specific Web site that shows the popularity.
- 3) The similarity of keywords as search terms entered by the user.

The output of the system is defined to include: Ranking the Each page, after pages are entered to the fuzzy system; internal calculations are performed on the system, then output with best ranks is created.

### 3. Modeling and evaluation

#### 3.1 Process

A summary of the process can be seen in Figure 1.

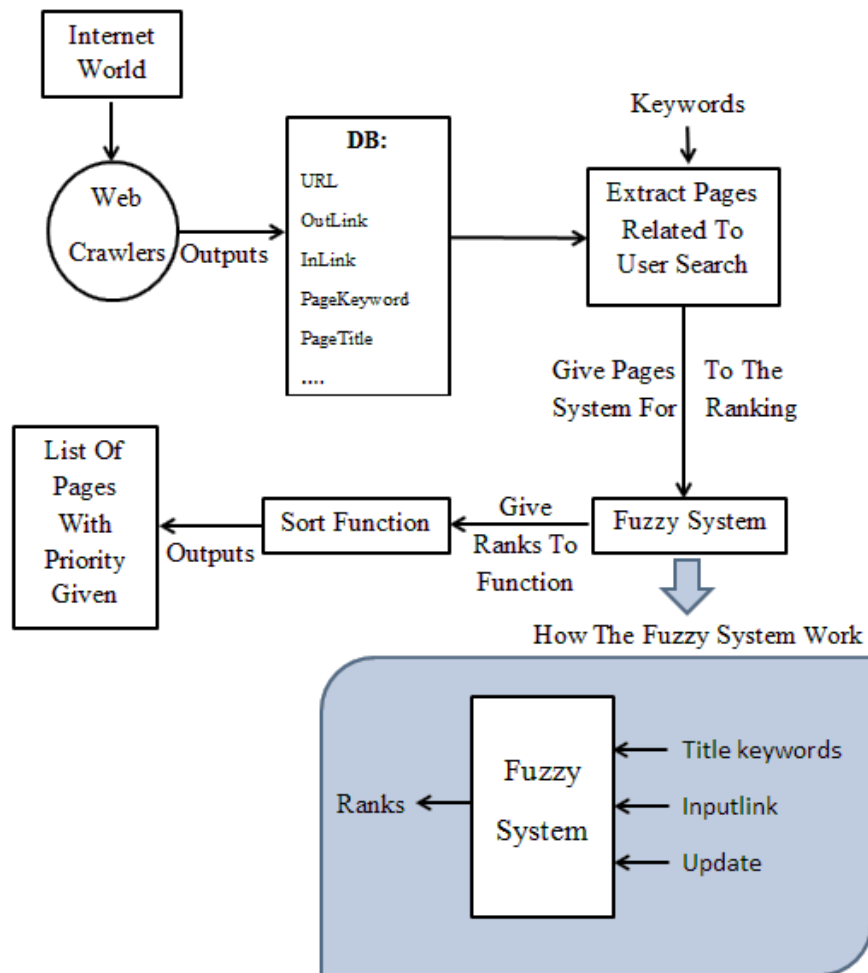


Figure 2: Description of work



---

**3.1.1 Web crawler:** The first step is to collect and create a useful database of several web crawlers on Persian web pages; then we can test the suggested work on this data.

**3.1.2 Extracting relevant pages:** After created a full database containing all the records for the project's requirements. Ranking criteria defined in the system to extract the appropriate output should be tested on the records. So when the user enters keywords to search the pages which is belongs, it's extracted from the database with the help of the keywords in the page and title. Then similarity between them is monitored; This action helps till least Pages and also relevant pages shown in the ranking for the fuzzy system.

**3.1.3 Ranking:** In this phase, pages mined from the earlier stage of the case had to be given to the fuzzy inference system and this system according to defined criteria and Ranges that is specified will determine the ranking which is appropriate.

**3.1.4 Ordering:** At the end, each page will rank according to the criteria and calculations on it by a fuzzy system. These are given to the sort function to rank the pages then we can see them as an output.

## 3.2 Fuzzy Inference System

This system, the main part of the suggested work is that with the database and sort function can greatly improve the ranking of the Persian materials. As explained earlier, the system includes an output and three inputs which will fully explain.

- The first input is the number of inbound links, which include the value between the numbers (0-1000) that zero shows the worst and 1000 shows the best. This value is determined by the average amount of inbound links to the Persian Web sites; it is





-----

possible that the higher number represents more popularity among the Board Persian language Members.

- The second input is the number of daily visits of Internet users from websites. The criterion for the amount of (0-4000) is defined. As previous when criterion is more shows that this web site is popular among users and they prefer to see the page.
- Last input is the similarity of search words with title, as explained before as a benchmark for measuring the similarity of page Compared to other pages within the database and into the system were considered. The basis of the values holds between (0-5). The higher degree of similarity shows the greater similarity page is marked with the word search. However, according to reviews the degree of similarity till two words is moderate.

All these criteria can be combined as good indication into one page to another's. This inputs only operates to achieve the best results in the ranking of these three criteria, have been reduced; Since the inception performance with inputs and other factors so further testing in the current system are able to be approved.

- After three criteria input system, their calculations, carried out on the pages, this work is to produce an output which is the rank of each page. For the single output value between (0-80) have been considered. As the user prefers to see the first two pages of search engine; namely maximum the first 20 pages returned; because he/she wants to save time and money; and actually there is no need to search more; because the best pages will be at the top rankings.

This project is based on the same three criteria for determining the best criteria to determine the many obsessions have concluded to help the user's require comparing search engines can be greatly relieved. The criteria for testing the accuracy of the system was run on

-----

a small database, still is able to get the desired results and satisfactory. Achieve good results in this project and this small database can we reached a conclusion of the project on the Internet database.

### 3.3 The rules

Fuzzy rules for implementing the system must be identified; but not any rules, these rules shall be determined with good accuracy to be ranked with the rules appropriate for that page. In this project, approximately 60 laws that fit the three criteria have been established. For all criteria, the membership functions of which consists of good and bad and very good and well defined that the lower, middle and best ranges shows on each criterion. Figure 3 is a diagram of the Rules that increase the number of visits of pages will lead to lower ranks; namely the initial pages of search engines is display.

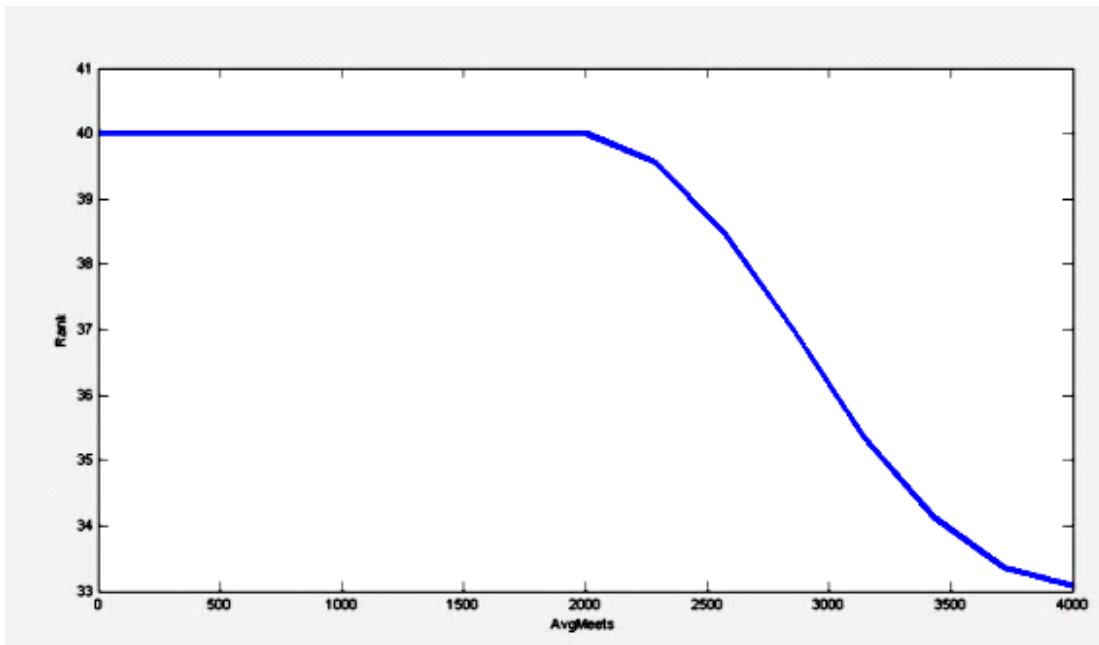


Figure 3: Two-dimensional chart with input averagemets



Finally, after building entrances, profile pages, and after evaluating the systems by laws, a number as a rank page is extracted. In the end all we have are the numbers to form a descending sorting function, sort and output to be displayed to the user.

#### 4. Testing

Finally, comparisons between the suggested works to other search engine were done. To do this, the two popular Webs search engine as Google and Yahoo and also four Persian search engines were used like parseek, parsjoo, Rismoon and jasjoo were the best engines. The first information that can be obtained from the comparisons, the number of pages returned by the search engine is to draw the graph of the first 50 pages are paid; according to Figure 2.

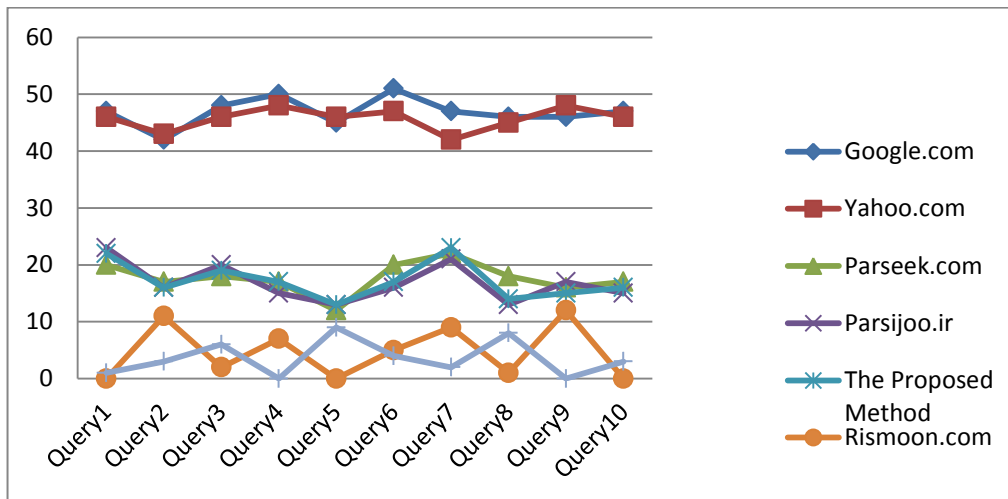


Figure 4: Number of relevant documents returned between the initial 50 documents

In figure 4 by entering 10 different search criteria in comparisons were made, where every time a different search term is entered. In many cases, the proposed solution is able to

work closely with other Persian search engine. Figure 5 accuracy of the proposed method for the top 10 results of search engines checks with different five queries. The diagram shows how the results are distributed on the list of top 10 results.

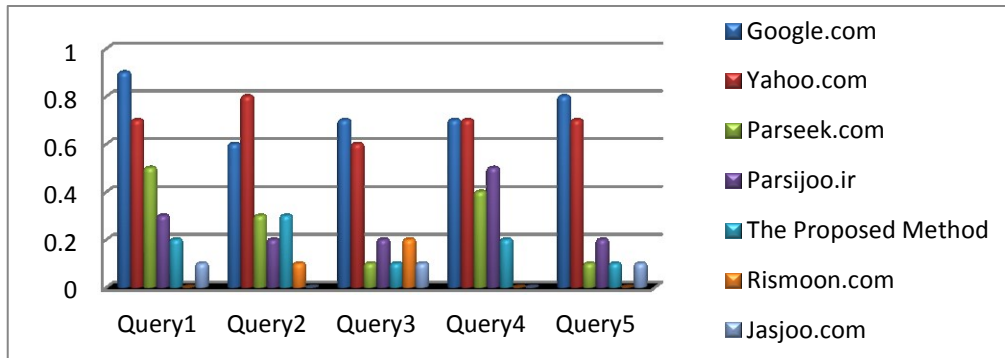


Figure 5: Chart of the Accuracy for top ten result

In figure 5 checked the accuracy average for this 7 method, which is derived from the insert 70 different search criteria.

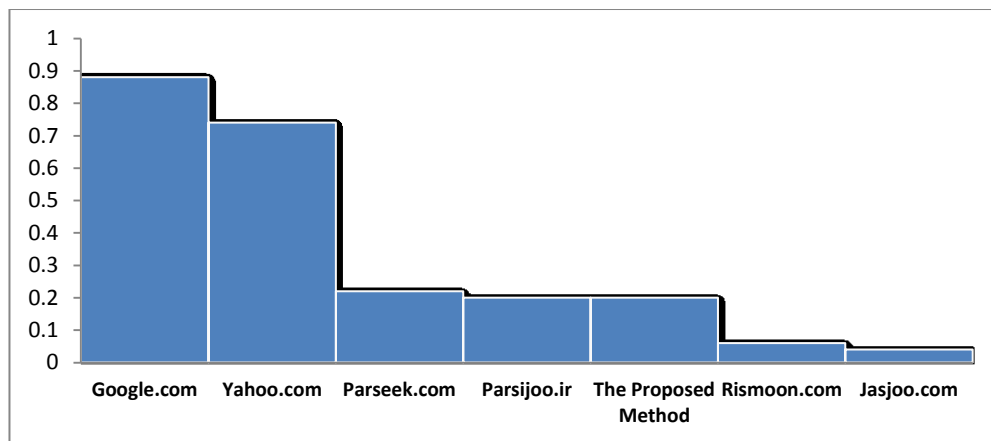


Figure 6: Chart of average accuracy for top 10 results with 70 search words

At the end we compared the response time of search engines using average the search words entered. As Figure 7; It is clear that a good proportion of our proposed method reached.

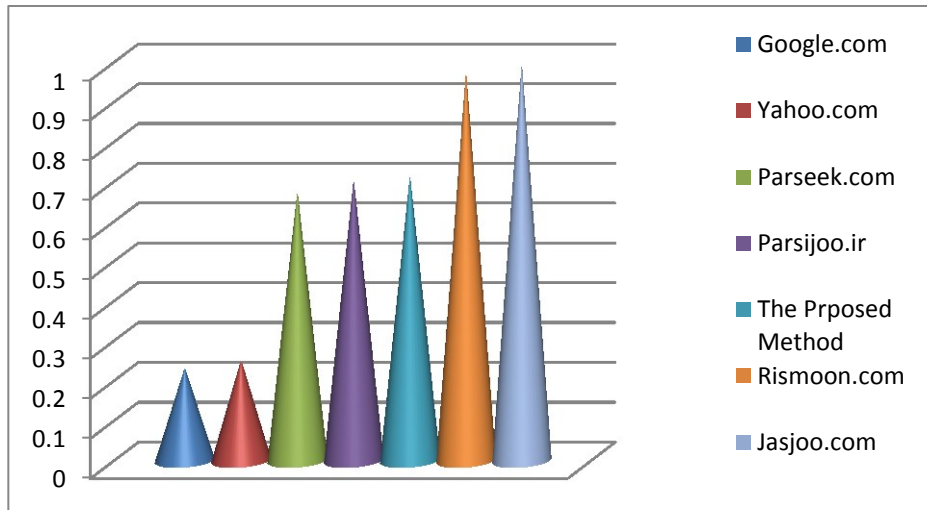


Fig.7: Chart of average response time by 20 search words

## Conclusions

Efficiency of the search engines which done the rankings of Farsi content was much lower than the English search engines. Only 70% of the results lead to the extraction of useful information. Although much research has been done on improving the search engine ranking of content; But since the focus is more on the subject English, so Persian users need to the best ranking Persian. In this paper, we first worked in the field of content ranking in the search engines and will review their progress; for reach this purpose, a better solution to improve the ranking was developed, which has a fuzzy inference system. The system is able



-----

to provide the best evidence that is intended for largely eliminate the need for Persian users and reach him to his pages. Finally, comparisons with other search engine also Persian search engines was done and the efficiency of the system was proved, since in many cases the results are closer to the user's search terms, as well as the results of other search engines on the Persian returns. The proposed method was also compared with the popular Web search engines and also to get a good comparison. In the future, our demand is to increasing the size of the database and according to it, increasing the criteria of the pages that are placed on the system input. To accomplish this, the system was used to lower size of the database; however, we reached good results and best ranks. Also by adding the entry criteria we should defined more rules so greater accuracy is required.

## **References**

- [1]. F. wang, yi li, y. zhang, An empirical "study on the search engine optimization technique and its outcomes, IEEE", 978-1-4577-0536-6/11/\$26.00 (2011).
- [2]. G. Poonkuzhali, r.k. Kumar, r.k. Keshav, k. Thiagarajan, K. Sarukesi, "Effective algorithms for improving the Performance of search engine results", IJOAMI, Issue 3, Vol. 5( 2011).
- [3]. S.A. Gollhofer, "Search engine ranking variables and algorithm"s, SEMJ.org, Vol. 1(august 2008).
- [4]. B.t. Sampath kumar, j.n. Prakash, "Precision and relative recall of search engines: a comparative study of google and yahoo", SINGAPORE JOURNAL of Library & Information Management, Vol. 38(2009).
- [5]. E. abdelraouf, N. L.y Badr, M.F. "Tolba, An efficient ranking module for an arabic search engine", IJCSNS, Vol.10 No.2 (February 2010).
- [6]. V. Ranjbar, A. Moghaddam, "Difference among ranking algorithms of different web search Tools: a statistical approach", Malaysian Journal, (2008).
- [7]. K. Taghva, R. Beckley, M. Sadeh, "A stemming algorithm for the farsi language", IEEE, 0-7695-2315-3/05 \$ 20.00(2005).
- [8]. R.L. Kumar, T. Gopalakrishnan, P. Sengottuvelan, "A Relational Based Fuzzy Clustering To Mine User Profiles for Web Directory Personalization", IJEAT, (2012).
- [9]. R. Rekik, I. Kallel, "Fuzz-Web: A Methodology Based on Fuzzy Logic for Assessing Web Sites", IJCISIM, (2013).



- 
- [10]. L.F. Lai, C.C. Wu, P.Y. Lin, "Developing a Fuzzy Search Engine Based on Fuzzy Ontology and Semantic Search", IEEE, (2011).
- [11]. A. Chaudhary, P. Punia, "OSA-PR: Optimized searching algorithm based on page ranking: Proposed algorithm", IEEE, (2012).
- [12]. K.C. Srikantaiah, P.L. Srikanth, V. Tejaswi, K. Shaila, K.R. Venugopal, L Patnaik, "Ranking Search Engine Result Pages based on Trustworthiness of Websites" , IJCSI, (2012).
- [13]. I.k. Shanna, N. Aggarwal, I. Duhan, R. Gupta, "web search result optimization by mining the search engine query logs", IEEE, 978-1-4244-9703-4/101\$26.00(2010).
- [14]. Z. huanjiong, "research a new method of search engine optimization", IEEE, 978-1-4244-5143-2/10/\$26.00, (2010).
- [15]. S.K. Ganta, S. Pavan, K. Somayajula, "Search engine optimization through Spanning Forest Generation Algorithm", IJCSE, Vol. 3 No. 9(2011).
- [16]. V. Kumar, G. Pooja, M. Kumari, A. Kumar, A. Appa rao, "Search engine optimization with Google", IJCSI, Vol. 9, Issue 1, No 3(January 2012).
- [17]. H.H. Kian, M. Zahedi, "An efficient approach for keyword selection; improving accessibility of web contents by general search engines", IJWEST, Vol.2, No.4(October 2011).
- [18]. H. Shahbazi, A. Mokhtaripour, M. Dalvi, B.T. Ladani, "A new approach for scoring relevant documents by applying a farsi stemming method in persian web search engines", Springer, CCIS 6(2008), pp. 745–748.
- [19]. X. Feifei, Z. Guangnian, "Design and implementation of a java-based search Engine algorithm analysis system", IEEE, 978-1-4244-3521-0/09/\$25.00(2009).
- [20]. A. Ramachandran, R. Sujatha, "Semantic search engine: a survey", IJCTA, (2011).