

# Search Engine Optimization by Fuzzy Classification and Prediction

<sup>1</sup>N . K. Senthil Kumar, <sup>2</sup>K. Kishore Kumar, <sup>3</sup>N. Rajkumar and <sup>4</sup>K. Amsavalli

<sup>1,2,3</sup>Department of Computer Science and Engineering, Vel Tech University, Chennai – 600062, Tamil Nadu, India; senthil.foss@gmail.com, kishorekumar.vnb@gmail.com, sivarajkumar.n@gmail.com

<sup>4</sup>Department of Computer Science and Engineering, Anand Institute of Higher Technology, Chennai - 603103, Tamil Nadu, India; amsa.gce@gmail.com

## Abstract

**Background/Objectives:** To optimize the search engine by fuzzy classification and prediction. **Methods/Statistical Analysis:** Accounting the characteristics such as page rank as usual, mouse movements of a user and the eye movements of a user while surfing a web page. The technique is proposed so that the search engine gets human like thinking and accordingly produces the search results. **Findings:** The proposed algorithm takes each and every parameter in user point of view and also the usual page ranking and applies fuzzy logic intelligence to predict the subsequent search results. Due to the right blend of the client qualities and web page attributes, the search engine will have the capacity to out-perform the current methods. **Application/Improvements:** Internet marketing and target sales

**Keywords:** Eye Movements, Fuzzy Logic, Mouse Movements, Page Ranking, Rational Search

## 1. Introduction

As a rule, a search engine comprises of crawler, indexer and ranker. A crawler retrieves web archives from the web<sup>1</sup>. Search engines make a map of the web by indexing web pages as indicated by keywords. The database of search engine preferably gives back a rundown of pertinent URLs, relating to the search keywords.

At first look, the administration search-engines appears to be exceptionally helpful and flawless, yet by a more cautious examination one can see the shortcomings in their search results. Among them are the high volume of pages and bandwidth confinements. Additionally in light of the fact that numerous web pages are oblivious internet, spider can't discover them<sup>2</sup>. Notwithstanding, elevated standards has made the occupation of search engine crawlers harder. For instance the crawler is required to search for and retrieve the pages containing the synonyms of the searched keyword as well. Since the two strings won't be comparable, every one of these

makes the occupation of a crawler much harder. In light of the fact that every dialect has its own particular special cases and search engine depends on string handling, perfect results can't be normal. Pages with a casual tone and slang or incorrectly spelled words add more troubles to the search-engines. The powerlessness of search-engines in handling parts of web pages like frames, Adobe Flash, pictures furthermore JavaScript and AJAX codes is an another shortcoming<sup>3</sup>. Notwithstanding ordinary difficulties specified, rivalry over webpage rankings has brought on a few webmasters to outline their websites and these wrongdoings include:

- Repeating words with a specific end goal to expand the thickness of the keyword searched.
- Exchanging random connections with websites that have a high ranking in the search engines.
- Generating dynamic pages with the motivation behind misleading search engines.

Concept-based IR is the search for data in view of its significance instead of the keywords searched<sup>4</sup>. This

\* Author for correspondence

methodology guarantees to build the nature of reactions since it catches the semantics of the archives. In this paper, an endeavor has been made for search-engines which will keep up client feedback, for example, mouse movements, eye movements rather than page investigation alone.

## 2. Fuzzy Sets and Fuzzy Logic

Utilizing fuzzy logic, a search engine can unwind the limits between word implications to a sure degree<sup>5</sup>. Pretty much as a camera lens set to a bigger aperture brings a more prominent range of perspective into center, fuzzy logic in a search engine will grow the profundity of field of potential search results.

Think of some as human expressions like “extremely adaptable”, “effectively coordinated” and “great arrangement”. This sort of dubious expressions is normal for the way we people impart through language and accordingly is an indispensable piece of our reasoning procedure. This stands out forcefully from the customary Boolean logic of computer programming which manages either true (1) or false (0) and nothing in the middle. Fuzzy logic overcomes any issues between them, giving a structure that permits you to numerically encode linguistic expressions and through that gives you an adaptable guideline based framework.

Basically, a fuzzy set is a set whose individuals from the set may have degrees of membership somewhere around 0 and 1, rather than traditional sets where every element must have either 0 or 1 as the membership degree - if 0, the element is totally outside the set; if 1, the element is totally in the set. As established logic depends on traditional set theory, fuzzy logic depends on fuzzy set theory.

To have the capacity to convey what needs be with this new fuzzy thing, we require some fundamental standards. Our definitive objective is to have the capacity to characterize logical expressions that we later can transform into articulations.

### 2.1 Linguistic Variables

A numerical variable takes numerical worth as Age = 65. A linguistic variable takes linguistic quality as Age = old. A linguistic quality says a fuzzy set.  $T(\text{age}) = \{\text{young, not youthful, exceptionally youthful, ... , moderately aged, not moderately aged, ... , old, not old, extremely old, ... pretty much old... .. etc}\}$ . Each of the worth for instance youthful, not youthful is translated as far as a membership capacity in fuzzy logic. Membership capacities can be characterized as triangular, trapezoidal, Gaussian and numerous types.

### 2.2 Fuzzy if-then Rules

Fuzzy is a principle based framework comprises of if-then standards, a cluster of certainties and an interpreter controlling the use of the rules given the truths. These if-then govern articulations are utilized to define the conditional explanations that comprise the complete learning base. A solitary if-then govern expect the structure ‘if x is A then y is B’ and the if-a portion of the standard ‘x is A’ is known as the antecedent or premise, while the then-some portion of the tenet ‘y is B’ is known as the consequent or conclusion. Some fuzzy standards are specified beneath.

If A and C then Y; If A and X then Z.

If B then X; If Z then D.

### 2.3 Fuzzy Operators

There are three important operators in the fuzzy set logic: OR and NOT. They likewise clearly exist for normal Boolean logic, yet we have to grow their definition to backing our new non-sharp membership capacities:

OR:  $A \cup B = \text{MAX}(A, B)$

AND:  $A \cap B = \text{MIN}(A, B)$

NOT:  $\neg A = 1 - A$

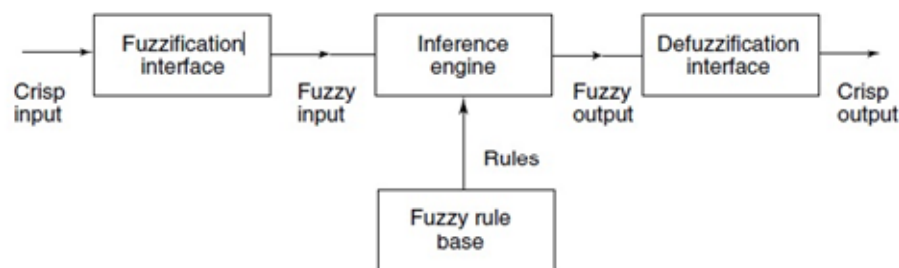


Figure 1. Fuzzy inference engine.

## 2.4 A fuzzy Engine

A fuzzy inference engine shown in Figure 1, first fuzzifies the fresh input. The inference engine apply the guidelines on the fuzzy input and create the fuzzy outputs. The outcomes are changed over to the fresh shape before submitted to the client.

## 3. Methodology

The proposed strategy means to give results which ought to mirror the human considering. For this reason, the input parameters from the people that is the clients of the web are thought seriously about and the search is enhanced by these parameters as the essential ones alongside the standard page ranking. The framework is isolated into the accompanying steps:

- Page Ranking.
- Mouse Movements.
- Eye Movements.
- Input parameters mapping.
- Decision making.
- Fuzzy Rules.

### 3.1 Page Ranking

PageRank is an algorithm utilized by Google Search to rank websites in their search engine results. PageRank was named after Larry Page, one of the authors of Google. PageRank is a method for measuring the significance of website pages. PageRank works by checking the number and nature of connections to a page to decide a harsh estimate of how imperative the website is. The hidden supposition is that more vital websites are prone to get more connections from different websites. If the number inbound connections to a web page, is all the more, then it is considered critical. Page ranking is additionally done taking into account various different variables too. Most critical of elements, for example, title, keywords, inbound connections are considered in this paper. Every one of the page ranking parameters considered are communicated as far as linguistic variables and the related linguistic qualities. A different fuzzy inference framework is utilized to discover the page rank mulling over the parameters as title, keywords, number of inbound connections.

For instance, the parameter title is a linguistic variable and its linguistic qualities are low, medium, high. As needs be, for each of the other two linguistic variables, linguistic qualities are allotted.

### 3.2 Mouse Movements Consideration

To overview the client conduct while managing undesired pages, a webpage may be outlined and the statistical society can be requested that locate the most intriguing philosophical passage inside of that page. Obviously, the page ought not simply loaded with philosophical passages and ought to contain four sections<sup>6</sup>. The first piece of the page incorporates various picture promotions that had no importance to the theme that the client was searching for and it is expected that the client would skirt this part quick. In the second section, a hamburger receipt ought to be portrayed. Clearly this part has nothing to do with what the clients will be searching for, either. As the part some time recently, it is normal that the clients would likewise skim this part rather rapidly, yet since the subject of this part is not conspicuous as effectively as the picture promotions, it is normal that clients would experience it with a to some degree lesser speed. The third part could be loaded with different passages however client shouldn't have the capacity to discover what he/she is searching for in this part either on the grounds that these passages contained raw numbers while clients where searching for the most intriguing philosophical passage. Last part was the place clients are required to invest more energy and proceed with more consideration and it might contain numerous philosophical passages. The said web page ought to contain a JavaScript code that, taking into account client conduct, caught vertical mouse speed, scroll speed, horizontal mouse speed, term time and standard deviation of horizontal mouse movement and passes them to a database. The parameters recorded for every four areas of the site should be independently computed and recorded in the database. An aggregate of no less than 50 computer understudies can take an interest in this test. As specified, members can be requested that pick the most intriguing passage inside of the page. Obviously clients won't know that their movements and conduct were recorded at the time. At the end of the day, clients will feel that they are partaking in an overview, in light of the fact that if they knew their movements were being recorded, it could influence their normal conduct.

A different fuzzy inference framework is utilized considering the mouse parameters mouse speed, horizontal movements and vertical movements and produces a fuzzy output. From the output of the framework, we can precisely predict the client conduct regarding the mouse movements.

### 3.3 Eye Movements Consideration

As partitioned fuzzy inference framework is utilized for discovering the eye movements of a human who surfs the web pages. Utilizing a web camera, the movements of eyes, for example, horizontal movements, vertical movements, to what extent a client sees a web page, whether the substance saw is a content substance or the substance is a picture content. Eye movements are critical that they really mirror the clients' perspective of the web page. At the point when a client surfs a web page for some substance, he may be occupied with pictures or he could be keen on content substance. It is select to discover where the enthusiasm of a client of a web page lies in. If the client likes to see pictures and invests more energy in taking a gander at pictures to get the data for which he is surfing the net, then the following search could concentrate on carry out searches with more picture based substance. Then again, if the client likes to take a gander at the content based substance as opposed to picture based ones, then the following search could concentrate on brining out pages with more content substance. Along these lines, we can decrease the download time for a client surfing the net as it is obvious that downloading a page with a bigger number of pictures sets aside additional time than a page with just content substance<sup>7</sup>. A fuzzy inference framework is utilized to figure out the eye qualities of a client and the output of this inference framework is supplied as input to an official choice making framework.

### 3.4 Parameters Mapping and Decision Making

If the search results are construct just in light of the page ranking procedures as utilized as a part of the greater part of the web programs in the business sector today<sup>8,9</sup>, then the clients' point of view and objectivity of search engine is lost. Then again, if the searching is construct just in light of the clients communication with the web pages and not on the on-page and off-page substance in the web pages, then the real concept of a search engine is vanquished. In this way, it is better if we consider the both of the parameters. Accordingly the fuzzified output of Page Ranking strategy, fuzzified output of mouse movements thought and the fuzzified output of eye movement thought are considered to a solitary fuzzy framework and produces a solitary output consolidating all the up to specified parameters. At the point when the client first searches, the page ranking system alone is utilized. Yet, as the client advances with

three or more pages, then the proposed algorithm works taking care of business and produces ideal result which would without a doubt upgrade the client's desire of a search engine.

### 3.5 Fuzzy Rules

#### 3.5.1 Fuzzy Rules for PageRanking

- If Inbound Links is high or title is low then PageRank is high.
- If Inbound Links is low or title is high then PageRank is medium.
- If title is high and keyword is high then PageRank is high.
- If title is low and keyword is low then PageRank is low.
- If title is medium and keyword is low then PageRank is low

#### 3.5.2 Fuzzy Rules for Mouse Movements

Consider Scroll Speed as SS, Vertical Mouse Speed as VMS, Horizontal Mouse Speed as HMS, Duration Time as DT, Mouse Movements as MM and Standard Deviation as SD for the following rules:

- IF (SS is very high) OR (VMS is very high).  
OR (SD is very low).  
OR (HMS is very high).  
OR (DT is very low).  
THEN (MM is very low).
- IF (SS is high) OR (VMS is high).  
OR (SD is low).  
OR (HMS is high).  
OR (DT is low).  
THEN (MM is very low).
- IF (SS is very low) OR (VMS is very low).  
OR (SD is high) .  
OR (HMS is very low).  
OR (DT is high).  
THEN (MM is high).

#### 3.5.3 Fuzzy Rules for Eye Movements

Consider waiting time as wt, Text Interest as TI, Text Content as TC and Image Content as IC for the following rules:

- If (wt is high) AND (TC is high) THEN TI is high.
- If (wt is high) AND (IC is high) THEN TI is high.

- If (wt is low) AND (TC is high) THEN II is high.
- If (wt is low) AND (IC is high) THEN TI is high.

### 3.5.4 Fuzzy Rules for the Search Decision Making System

Consider Page Ranking as PR, Mouse Movements as MM for the following rules:

- If (PR is low) AND (MM is high) AND (TI is high) THEN Interest is high.
- If (PR is low) AND (MM is low) AND (TI is high) THEN TI is medium.
- If (PR is high) AND (MM is medium) AND (TI is high) THEN Interest is high.
- If (PR is low) AND (MM is medium) AND (II is medium) THEN Interest is medium.
- If (PR is medium) AND (MM is high) AND (II is low) THEN Interest is medium.

## 4. Conclusion

The proposed search engine improvement procedure can be actualized considering the qualities, for example, page rank of course, mouse movements of a client and the eye movements of a client while surfing a web page. Due to the right blend of the client qualities and web page attributes, the search engine will have the capacity to outperform the current ones. The proposed procedure can be tried amongst a gathering of no less than 100 individuals to test its exactness and the in future more client qualities can be considered to further enhance the precision of the fuzzy inference framework.

## 5. References

1. Pinkerton B. Finding what people want: Experiences with the webcrawler. USA, University of Wasington: 2000.
2. Huillier GL, Alvarez H, Ríos SA, Aguilera F. Topic-based social network analysis for virtual communities of interests in the dark web. ACM SIGKDD Workshop on Intelligence and Security Informatics. 2010 Dec; 12(2):66–73.
3. Baeza-Yates R, Broder AZ, Maarek Y. The new frontier of web search technology: Seven challenges. Search Computing II, Springer-Verlag: Berlin Heidelberg. 2011; 6585:3–9.
4. Haav HM, Lubi TL. A survey of concept-based information retrieval tools on the web. 5th East-European Conference ADBIS; 2001. p. 29–41.
5. Sherman C. In praise of fuzzy searching. 2001. Available from:
6. <http://searchenginewatch.com/sew/news/2064695/in-praise-fuzzy-searching>
7. Huang J, White RW, Dumais S. No clicks, no problem: Using cursor movements to understand and improve search. Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems. 2011. p. 1225–34.
8. Cooke L. Is the mouse a “Poor man’s Eye Tracker”? Annual Conference-Society for Technical Communication; 2006. 53. p. 252.
9. Koteeswaran S, Visu P, Kannan E. Enhancing JS–MR based data visualisation using YARN. Indian Journal of Science and Technology. 2015 Jun; 8(11):1–5.
10. Mohammadi R, Ghaffari A. Optimizing reliability through network coding in wireless multimedia sensor networks. Indian Journal of Science and Technology, 2015 May; 8(9):834–41.