A Major Project Report On

METAFUSION: AN EFFICIENT METASEARCH ENGINE USING GENETIC ALGORITHM

Submitted in partial fulfillment of the requirements
for the award of the degree of

MASTER OF TECHNOLOGY

IN

SOFTWARE ENGINEERING

By

Devika Singh

(Roll No. 2K14/SWE/04)

Under the guidance Of

Dr.Daya Gupta

Professor

Department of Computer Science Engineering

Delhi Technological University , Delhi

Department of Computer Science Engineering

Delhi Technological University, Delhi

2014-2016

DELHI TECHNOLOGICAL UNIVERSITY

CERTIFICATE

This is to certify that the project report entitled **METAFUSION: AN EFFICIENT METASEARCH ENGINE USING GENETIC ALGORITHM** is a bona fide record of work carried out by Devika Singh(2K14/SWE/04) under my guidance and supervision , during the academic session 2014-2016 in partial fulfillment of the requirement for the degree of Master of Technology in Software Engineering from Delhi Technological University, Delhi.

To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University /Institute for the award of any Degree or Diploma.

Dr. Daya Gupta

Professor

Department of Computer Science Engineering

Delhi Technological University, Delhi

DELHI TECHNOLOGICAL UNIVERSITY

ACKNOWLEDGEMENTS

I feel immense pleasure to express my heartfelt gratitude to Dr. Daya Gupta for her constant and consistent inspiring guidance and utmost co-operation at every stage which culminated in successful completion of my research work.

I also would like to thank the faculty of Computer Science Engineering Department, DTU and my peers for their kind advice and help from time to time.

I owe my profound gratitude to my family which has been a constant source of inspiration and support.

Devika Singh

Roll No. 2K14/SWE/04

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

World Wide Web is a dynamic source of information which is expanding its content at a staggering rate. Individual search engines are not able to handle the exponential nature of web. Hence meta-search engines are used to solve the problem of low web space information coverage rate of individual search engines. A meta-search engine is a kind of search tool that dynamically dispatches user query to the underlying search engines, hence providing parallel access to multiple search engines and then aggregate the results to present single consolidated result list to user. In this research work , a novel meta-search engine, *MetaFusion* , has been proposed. The proposed algorithm uses Fuzzy AHP along with Genetic algorithm to get more comprehensive and optimized results. Fuzzy Analytical Hierarchy Process reflects human thinking and addresses the uncertainty of information while making decisions in MCDM problems. Genetic Algorithm(GA) is highly robust and self- adaptive algorithm , hence, solves the more complex problems in optimum manner.GA uses average weight of document in underlying search engine as fitness function for merging results. Experimental results shows that relevancy of returned results by *MetaFusion* is more than several existing research Metasearch engines. The precision of proposed model *MetaFusion* comes out to be more when compared with available research metasearch engines i.e. Dogpile ,Infospace and PolyMeta.

***Keywords**:* MetaFusion, Information retrieval, metasearch, search engine, MCDM, Fuzzy AHP,Genetic Algorithm

CHAPTER 1: INTRODUCTION

1.1 Overview

These days common man are using World Wide Web to search needed information using variety of Search engines like Google ,Bing , Yahoo, Ask , Lycos etc. World Wide Web is a huge repository of data consisting of billions of web documents that are distributed over multiple web servers. The information on web is increasing day by day and due to this web coverage given by individual search engines have been constantly decreasing[1]. The web contains largely unregulated documents whose contents are updated regularly. Therefore , degree of documents content quality and reliability varies with time. For individual search engines only return 45% of relevant results[2]. The research in last decade has focused on improving the search procedure. As a result meta-search engines have been proposed which greatly improve search results. Several research studies have demonstrated that meta-search engines can upsurge the search effectiveness significantly [3,4,5,6]. Various research meta-search engines are available online like Dogpile , Infospace , Excite, Polymeta, DuckDuckGo etc. which are used by users to serve their needs.

1.2 Basic Concepts

1.2.1 Meta-search Engine Concepts:

A meta-search engine is basically considered to be a fusion tool which commences its session when user poses query to its interface. After that MSE processes the query and submits the refined query to multiple underlying search engines. The underlying search engines accesses the network resources and then return back their respective results to MSE. Then MSE aggregates the returned results into single consolidated rank list by using certain aggregation algorithm.

The basic functioning of meta-search engine is shown in Figure 1. [7]

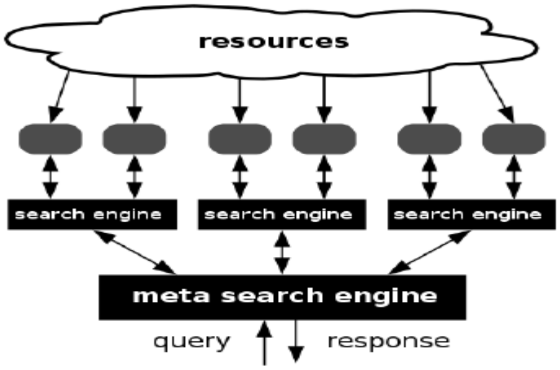


Figure 1: Meta-search Engine Functioning [7]

Meta-search engines don’t have their own file index database , hence, they forward the query to several underlying search engines, and after that merges their respective results using certain aggregation algorithm. Hence Query dispatching and Result aggregation are the two major functions of meta-search engine [8]. Therefore, meta-search engine can be considered as an interface on the top of multiple search engines to provide the user with uniform access to many search engines at once.

1.2.2 Meta-search Engine architecture

The basic architecture of meta-search engine is shown in Figure 2. The following tasks are performed by MSEs sequentially:

i. Accepts user query.

ii. Pre- Processes the submitted query.

iii. Passes query to the underlying search engines.

iv. Combines search results of different search engines using certain aggregation algorithm to generate single consolidated rank list.

v. Performs post-processing on returned results and displays it to the user.

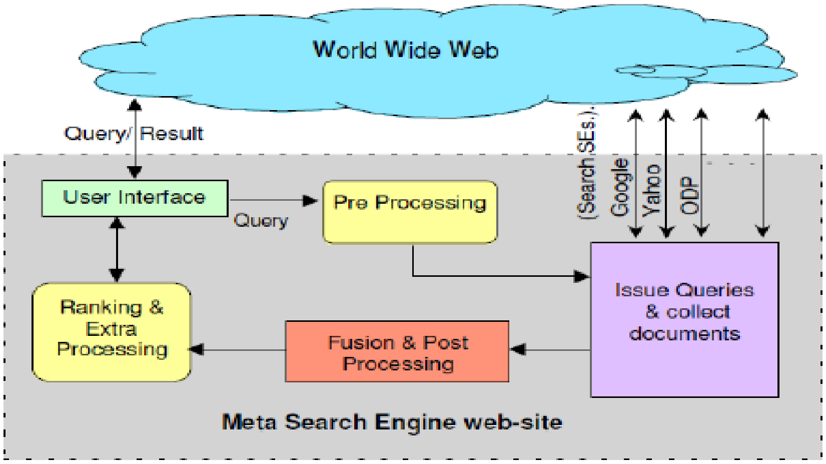


Figure 2: Meta-search Engine Architecture[9]

1.2.3 Advantages of Meta-Search Engine

The Meta-search engine is an improvement over a single search engine since it broadens the search coverage and hence , allows the extraction of more appropriate results with the same amount of effort. The advantages of meta-search engine in information retrieval can be concluded as follows [10]:

i. Increases search effectiveness by increasing web search coverage.
ii. Improves search accuracy.
iii. Increases users convenience by allowing him to access multiple search engines for just one query.
iv. Solves real time search issues related to web search.
v. Improves retrieval results by invoking multiple search engines in parallel.
vi. Addressing the scalability of searching the entire Web.

1.3 Motivation

World Wide Web has become the main place for searching information on any topic. This makes searching a key activity and thus , search engines the most widely used tools on the Web. The research in last decade has focused on improving the search procedure. As a result meta-search engines have been proposed.

Early meta-search engine models MetaCrawler[11,12] , Borda-Fuse[13] were centered around assigning weight scores to documents , and not considering search engine's importance weights. Then came next generation of MSE like Weighted Borda Fuse [13] where individual search engines are also assigned weights to reflect their performance. Recent models of MSE [14] were based on OWA and use multi –criteria decision making. It addresses the issues related to missing documents. OWA operator provides efficient method to merge results. Another recent model of MSE , MetaSurfer [15], uses modified EOWA along with FAHP to perform meta-search. Very recent model of MetaXplorer[16] is based on Intelligent OWA operator along with Fuzzy Analytical Hierarchy Process to evaluate document score.

But the earlier Meta-search engine models were unable to handle dynamic nature of web .Search engines performance varies with time because of updation of ranking algorithm , modification in indexing and updation of database happens. *Therefore , we need a Meta-search engine model which adapts with the changing needs of environment.*

Also the Ordered Weighted Averaging operator used in MCDM problems till now are dependent on the decision maker's judgments . Experts assigns importance degrees to search engines i.e. criteria and thus making a process biased. *Therefore, we need a model in which importance degree to search engine's are assigned in unbiased manner.*

Earlier developed meta-search engine models only perform pre-processing and post-processing of results. No internal processing was performed on returned results. This motivated us to propose a meta-search model which performs internal processing on returned results.

Furthermore, the previously existing models have not combined Fuzzy AHP with Genetic Algorithm to retrieve results in web information retrieval domain. Therefore , we need a model which applies multi-optimization algorithm in result merging to generate optimal results.

The performance of meta-search engine models is evaluated by using precision as a metric. Very recent model of meta-search[16] evaluated its performance by considering 30 queries . But this evaluation should consider more number of queries to improve precision metric.

From the above discussion , it can be clearly noted that we need an adaptable meta-search engine model which changes with changing need of environment. Also , importance weights to search engines should be assigned in unbiased manner. Furthermore, there is a need to apply multi-optimization algorithm to generate optimal results. This motivated us to pursue the research in the field of meta-search engine so that we can address the problems of existing MSEs. Also evaluation should be improved by considering more number of research queries.

1.4 Problem Statement

In this research we are improving forgoing MSE by applying multi optimization algorithm. This research work present a new meta-search engine, named as MetaFusion, which is capable of handling dynamic web environment. The proposed algorithm uses Fuzzy AHP along with Genetic algorithm to retrieve comprehensive results. FAHP reflects human thinking and addresses the uncertainty of information while making decisions in MCDM problems. Genetic Algorithm is a self adapting global optimization parallel search algorithm which imitates biological evolution process i.e. crossover ,mutation, selection [17].The main challenge during information retrieval is to find most appropriate set of documents with respect to user query. Hence Genetic Algorithm(GA) is used for result merging which uses average weight of document in underlying search engine as fitness function. The documents are ranked in decreasing order of their fitness value i.e. most relevant document have higher fitness value and is present at top position in rank list. In-OWA operator is used in training phase to assign importance degree to search engines. To reflect the environmental changes, training algorithm should be run on a periodic basis. This makes our proposed model, MetaFusion , adaptable with changing needs. Furthermore, URL analysis is also performed on returned results to incorporate the measure of internal processing.

Hence the problem of this thesis can be stated as:

"Proposing an Adaptable and Efficient Meta-search engine model, MetaFusion , which uses Fuzzy AHP along with Genetic Algorithm to generate single consolidated rank list of results returned by individual search engines."

1.5 Scope of Work

The performance of our proposed model, MetaFusion , is evaluated by considering 100 test queries from different domains of real world. The performance of MetaFusion is compared with existing research MSEs i.e. Dogpile , InfoSpace and Polymeta in terms of precision.

The scope of work can be summarized as:

i. Designing the user interface of MetaFusion to accept user query and then dispatch it to several underlying search engines.

ii. In –OWA operator is used to assign importance degree weight to search engines. Hence makes our process free form judgments of decision maker's.

iii. URL analysis is performed to analyze the returned documents relevance.

iv. Fuzzy AHP is applied to address the uncertainty factor associated in Multi-Criteria Decision Making Problem.

v. Genetic Algorithm is applied to merge the result and form the single consolidated rank list of documents based on the fitness value of document. In the proposed work OWA operator have been used as Fitness function.

vi. The performance of proposed model, MetaFusion, is evaluated in terms of precision by considering a set of 100 test queries taken from different domains of real world.

vii. Then we have compared the obtained precision value of MetaFusion with popular existing research MSEs i.e. Dogpile, Infospace and Polymeta.

1.6. Thesis Organisation

Further thesis is organized as follows:

Chapter 2 : This chapter presents the literature review of existing metasearch models. Also comparative analysis between the models is described.

Chapter 3 : This chapter presents the detailed description of proposed metasearch model, MetaFusion.

Chapter 4: This chapter describes the implementation deatils of this research work.

Chapter 5 : This chapter presents the evaluation of the proposed MSE, MetaFusion. It also compares the performance of MetaFusion with three popular MSEs i.e. Dogpile, InfoSpace and PolyMeta.

Chapter 6: This chapter concludes the thesis and discusses the possible improvements in this research work in future.

Chapter 7 : This chapter deals with publications from this research work.

CHAPTER 2: LITERATURE REVIEW

In this chapter we describe the literature survey of previously existing meta-search engine models . At the end of chapter comparative analysis between these models is also presented.

2.1 MetaCrawler

MetaCrawler is one of the first meta-search engines developed by Erik Selberg and Oren Etzioni at the University of Washington, Seattle in 1995[11,12].

The steps involved in ranking computation is described below:

1. User query is processed and forwarded to underlying search engine's such as Lycos, Excite, Yahoo etc.
2. The documents are assigned weights i.e "confidence score" in range of 0 to 1000 such that top most document in each search engine's result list gets highest value of confidence score.
3. Then results are merged by adding corresponding values of confidence scores.
4. Finally the duplicates are removed and result is displayed to user.

The control flow of MetaCrawler is shown in Figure 3.

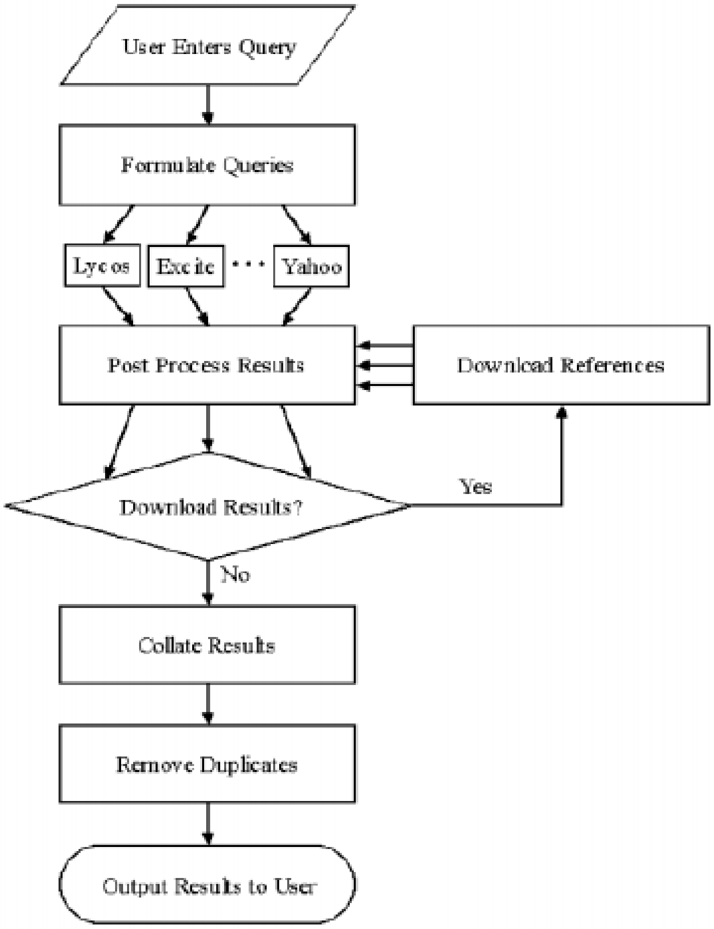


Figure 3: Control Flow of MetaCrawler[38]

2.1.1 Model Evaluation

Precision and recall are the standard criteria's used for evaluating meta-search models.E. Jacob and Manoj [9] studied and demonstrated that MetaCrawler precision value comes out to be 0.35 when we consider top 20 documents returned over 12 independent test queries. Also Alexa[18] Internet web service which is subsidiary of Amazon.com[9] published MetaCrawler as third most popular meta-search engine. Alexa Toolbar records user's visit count as a metric to rank its website. Also, number of in-linking pages to meta-search engine is considered as an another metric to measure MSE's popularity. By considering in-linking factor as metric, MetaCrawler was ranked second according to Google and sixth according to Yahoo[9].

2.2 Borda-Fuse Model

Borda-Fuse Model was proposed by Aslam and Montague [13] in 2001 for result aggregation.

The steps involved in ranking computation is described below:

1. Each search engine is considered as a voter , which ranks a set of *N* documents according to their relevance.
2. The top most document is assigned N points , the second one is assigned N-1 points and so on procedure continues.
3. The documents that are missing in search engine result list are assigned remaining points evenly.
4. Then for each document , we add the corresponding Borda Points obtained from different search engines .
5. Display the result in decreasing order of Borda Point value. Hence top most document has highest Borda Points.

2.2.1Model Evaluation

Borda-Fuse meta-search model performance is evaluated by Aslam and Montague[13] , using datasets offered by (Text retrieval conference)TREC 3, TREC 5, TREC 9 and Vogt dataset. Each TREC dataset consists of 50 queries and Vogt Dataset[19] consists of 10 queries. Precision value is used to evaluate and compare Borda-Fuse model performance with other meta-search models which is shown in Figure 4.

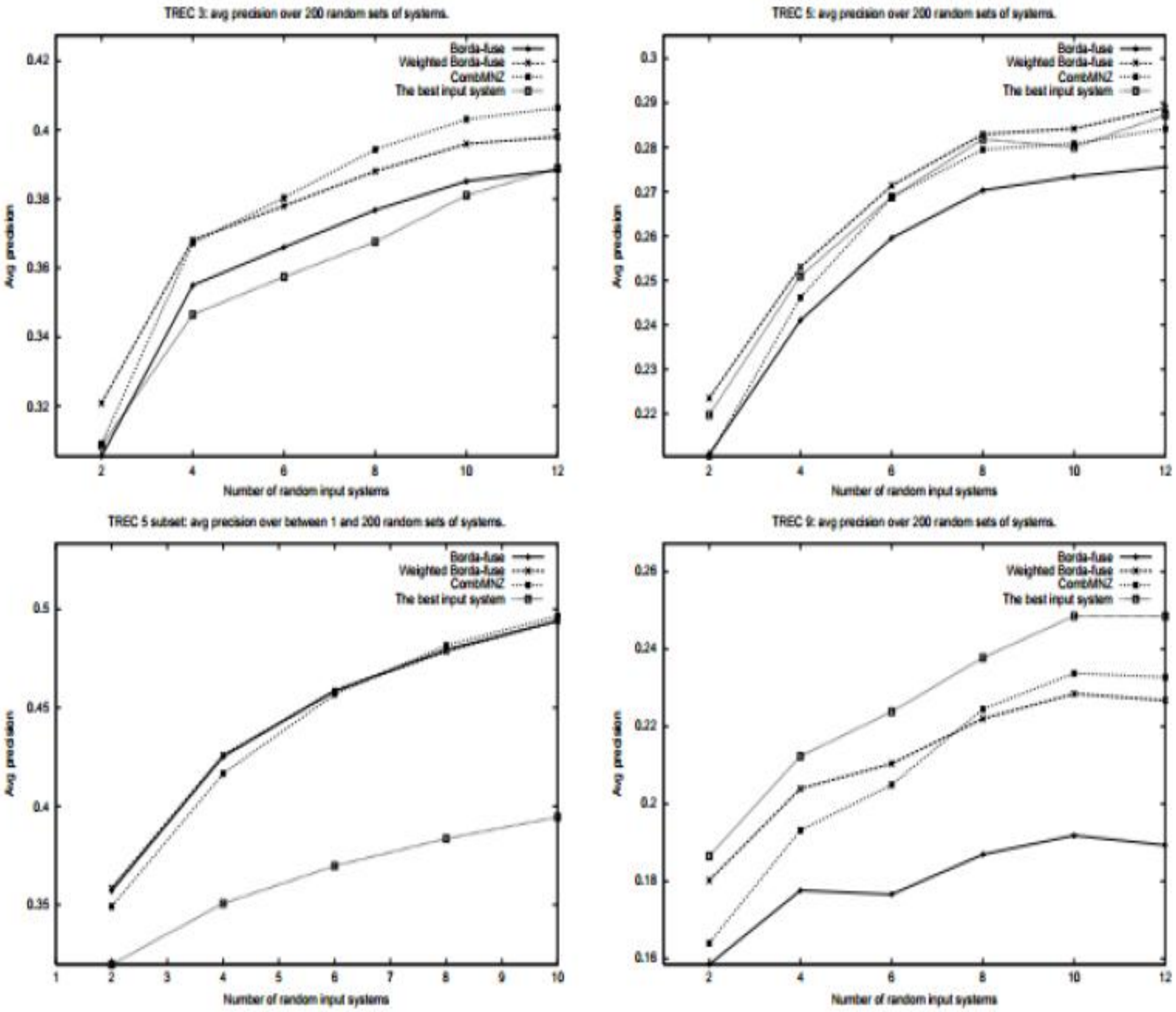


Figure 4: Weighted Borda Fuse and Borda Fuse Evaluation[13]

2.3 Weighted Borda-Fuse Model

The major drawback of Borda-Fuse Model was that it considers search engine selection process to be homogeneous i.e. all search engines are assigned equal importance degree weights . But in real world this situation does not happen. Each individual search engine should be assigned different importance degree weight to reflect their performance. Therefore , Weighted Borda Fuse model[13] was proposed to incorporate heterogeneous nature of web.

The steps involved in ranking is described below:

1) User posted query is preprocessed and forwarded to multiple search engines.
2) Each individual search engine result list contains N number of documents where top most document is assigned N points and the document at second position is assigned N-1 points . The process continues until all documents are assigned points.
3) The documents that are missing in search engine result list are assigned remaining points evenly corresponding to that search engine.
4) In next step, the Borda Points are multiplied with search engine's importance weights.
5) Then we add corresponding Borda points for each document present in different search engine's result list to generate total Borda Points for each document. Therefore, total Borda points are considered as weighted sum of Borda-points assigned by different search engines.
6) In last step, final aggregated rank list is formed by merging results of individual search engine's . The documents are arranged in decreasing value of Borda points , hence, the topmost document should be having highest value of Borda point.

2.3.1 Model Evaluation

Weighted Borda-Fuse meta-search model performance is evaluated by Aslam and Montague[13] , using datasets offered by (Text retrieval conference)TREC 3, TREC 5, TREC 9 and Vogt dataset. Each TREC dataset consists of 50 queries and Vogt Dataset[19] consists of 10 queries. Precision value is used to evaluate and compare Weighted Borda-Fuse model performance with multiple other meta-search models which is shown in Figure 4. It can be clearly observed that Weighted Borda Fuse performance is better when compared with Borda – Fuse model.

2.4 OWA Model

Earlier discussed models like Borda-Fuse and Weighted Borda-Fuse does not handle missing documents properly. They generally assign remaining points evenly amongst missing documents , thereby, causing them to be present at the bottom position in result list. But , if a document is missing in one search engine's result list and present in another search engine's list , it does not makes it less relevant. Due to large domain of web space, it is not possible for each search engine to cover entire portion of web space. Hence documents appear missing in one search engine's result list but present in another. Therefore , OWA model was proposed by Diaz et al.[14] to address this issue.

2.4.1 Ordered Weighted Averaging Operator Concepts

There are two extremes that are defined by the classical binary logic i.e.

i. "or" where atleast one of the criteria should be met .

ii. "and" where all the criteria should be met.

In 1988, Yager[20] proposed **Ordered Weighted Averaging (** OWA) operator which is applied in Multi Criteria Decision Making Problem to aggregate scores. Therefore OWA operator is used in making overall decision.

OWA operator of dimension n is defined as a function F: $R^n \rightarrow R$(where R = [0,1]), with associated weighing vector W , where W=[W_1 W_2 W_3….W_N] , such that

i. $W_i \in [0,1]$

ii. $\Sigma W_i = 1$

iii. $F(A_1,A_2,A_3 . . .. ,A_n) = W_1*B_1 + W_2*B_2 +W_3 *B_3 ..… + W_n*B_n$

where Bi is the i^{th} largest value in A_1, A_2, ...,A_n.

Consider an example where F is an OWA operator with dimension n=4 is given as F=[0.2,0.7,0.4,1.0] and let W =[0.2, 0.4, 0.15 ,0.25] be an associated weighting vector.
Hence , ordered argument vector B is given as [1.0,0.7,0.4,0.2] which is formed by rearranging F values in decreasing order.

F(0.2,0.7,0.4,1.0) = W*B

=$[W_1*B_1+W_2*B_2+W_3*B_3+W_4*B_4]$

= (0.2)(1) + (0.4)(0.7) + (0.15)(0.4) + (0.25)(0.2)

= 0.59

There are various applications of OWA operator in real life applied in MCDM field which are listed below:

- Doctoral Student Selection problem [21]
- Data modeling and re-identification in Data Mining [22]
- Applying OWA operator in Minkowski distance [23]
- Sports Management [24]
- Meta-search based information retrieval [25]

2.4.2 OWA operators Evolution

OWA operator has evolved with time. Chiclana[26] introduced Ordered Weighted Geometric (OWG) operator in 2000 for ratio-scale measurements because geometric mean is better suited for ratio-scale measurements [27,28] as compared to arithmetic mean .Induced OWA (IOWA) operator was proposed by Chiclana[29] in 2007, which introduces the concept of order inducing variable for reordering the arguments. Importance Induced OWA operator (I-IOWA) assigns different important degrees to criteria and hence reorders the arguments on the basis of criteria importance degrees.

Consistency IOWA operator (C-IOWA) uses consistency index value of the experts to perform argument reordering .Since OWA operator takes numerical values as input , hence, it is not able to handle the linguistic data. Zarghami[30] proposed EOWA operator, which incorporates the concept of linguistic inputs. Linguistic inputs are represented by their equivalent triangular fuzzy numbers. After that crisp numbers are obtained by using the max-membership method. EOWA operator cannot handle the uncertain inputs whose values are known only under pessimistic and optimistic conditions. Hence, Suo [31] introduced Advanced OWA (AOWA) operator in 2012, which uses the concept of interval theory to represent uncertain arguments and applied Center Of Gravity (COG) method for defuzzying.

2.4.3 Application of OWA in MCDM

OWA operator is generally applied in meta-search engines for result aggregation [14,32]. The steps involved are listed below:

1) Assign 'Positional value' to each document present in result list.
2) Positional value of document d_i in result list rl_j returned by search engine s_j is calculated by using following formula:

$$PV= (n- r_{ij}+1)$$

where r_{ij} = rank of document d_i in search engine $s_{j,}$

n= total number of documents present in the result.

3) Therefore, documents present at top position in result list will be having higher positional value.
4) Hence, Positional Values are a measure of the degree to which a document (analogous to a MCDM 'alternative') satisfies a search engine's (analogous to MCDM 'criteria') criteria for retrieval.
5) OWA model has proposed two heuristics i.e. H_1 and H_2 to appropriately handle missing documents. Heuristic H_1 equation is listed below:

$$PV= \sum_{i=1}^{r} \frac{PVi}{m}$$

Therefore in H_1 heuristic , missing document positional value is calculated by taking average of positional value in m search engines where it appears.

Heuristic H_2 equation is listed below:

$$PV= \sum_{i=1}^{r} \frac{PVi}{k}$$

Therefore in H_2 heuristic , missing document positional value is calculated by taking average of positional value in k search engines where it appears.

6) Calculate OWA operator weight using following equation:

$$W_i = \left(\frac{i}{n}\right)^{\propto} - \left(\frac{i-1}{n}\right)^{\propto}$$

Where n is number of criteria i.e. search engines and α ε[0,1].

7) In next step , OWA operator mapping function(F) is evaluated by giving input to F as OWA weights W_i and Positional Value (PV_i) of documents.

$$F(d)= \sum_{j=1}^{m} W_j * PV_j$$

8) Finally in last step documents are ranked in decreasing order of function F value.

2.4.4 Model Evaluation

OWA model performance is evaluated by Diaz et al.[14] , using datasets offered by TREC(Text retrieval conference). The dataset contains 50 test queries and 40 search systems. Diaz considered precision value as a metric to evaluate and compare OWA model performance with Borda Fuse model which is shown in Figure 5. Tests were performed by considering different quantifier values (α) i.e. 0.5 , 1 , 2 and 2.5. It can be clearly observed that OWA model performance is better when compared with Borda –Fuse model in terms of precision.

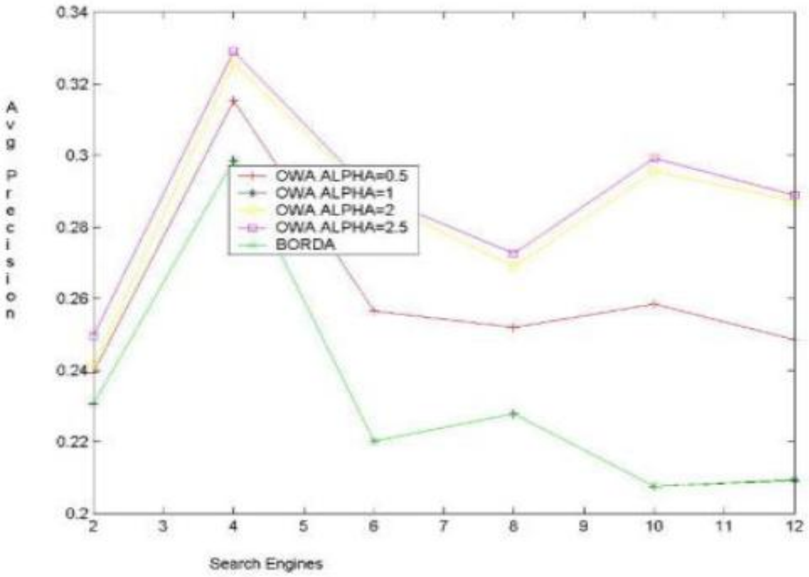


Figure 5: Evaluation of OWA Model [14]

2.5 Fuzzy Analytical Hierarchy Process

Satty[33] proposed Analytical Hierarchy Process (AHP) in 1980, which is used in making decisions to solve MCDM problems and is based on pair wise comparison on ratio scale. Satty[34] demonstrated the benefits of applying pair wise comparison in MCDM problems. AHP basically incorporates human thinking in making sound decisions about small problems. But , AHP can't handle the uncertainty and imprecision associated with the decision maker's perception.

Therefore in 1988 Fuzzy Analytical Hierarchy Process(FAHP) was introduced which reflects human thinking while making decisions[35]. Fuzzy AHP uses linguistic quantifiers while making comparisons instead of crisp numbers . Hence, crisp judgments gets transformed into fuzzy judgments.

FuzzyAHP steps is applied in the following manner:

1) A *NxN* matrix is constructed where *N* denotes the number of alternatives. Each entry in the matrix is a linguistic variable i.e. less Important, Important, more important, etc. which is represented by triangular fuzzy numbers.

$$\check{A} = \left(\check{a}_{ij}\right)_{nxn} = \begin{bmatrix} (1,1,1) & (l_{12}, m_{12}, u_{12}) & \dots & (l_{1n}, m_{1n}, u_{1n}) \\ (l_{21}, m_{21}, u_{21}) & (1,1,1) & \dots & (l_{2n}, m_{2n}, u_{2n}) \\ \dots & \dots & \dots & \dots \\ (l_{n1}, m_{n1}, u_{n1}) & (l_{n2}, m_{n2}, u_{n2}) & \dots & (1,1,1) \end{bmatrix}$$

2) Convert triangular fuzzy numbers into fuzzy interval by using α- cut based method .

$$\alpha\text{left} = [\alpha^* (m\text{-}l)] + l \ , \quad \alpha\text{right} = u\text{-} [\alpha^*(u\text{–}m)]$$

where (l,m,u) represents triangular fuzzy number and $\alpha \, \varepsilon \, [0,1]$ denotes confidence factor.

$$\check{p}_{\alpha} = \begin{bmatrix} (\alpha\text{left}_1 \alpha\text{right}_1) \\ (\alpha\text{left}_2 \alpha\text{right}_2) \\ \dots \\ (\alpha\text{left}_n \alpha\text{right}_n) \end{bmatrix}$$

3) Then, crisp value of one alternative over every other alternative is represented by Crisp Judgment Matrix i.e. C_λ by using following formula:

$$C_\lambda = \lambda^*\alpha\text{right}+(1\text{–}\lambda)^*\alpha\text{left}$$

where λ which is known as Optimism Index of Decision Maker lies between[0,1]

$$C_{\boldsymbol{\lambda}} = \begin{bmatrix} C_{\lambda 1} \\ C_{\lambda 2} \\ \dots \\ C_{\lambda n} \end{bmatrix}$$

2.5.1 Fuzzy AHP Application Areas

- Remote Sensed Data [36]
- Evaluation and Selection of Construction Project Contractor [37]
- E-commerce success factors evaluation [38]
- Capital Investment [39]
- GIS Application [29]
- Project Risk Assessment [40]
- Evaluation Of Green Products Design [41]

2.6 Hybrid Fuzzy Model

The major drawback of OWA model is that it does not consider user preferences or choice for search engines selection during performing aggregation in MCDM problems. Also ,OWA model ignores the correlation relationship between search engines and documents that may influence the search quality. Hence , Hybrid Fuzzy model was proposed by De and Diaz[42] in 2009 to overcome this limitation.

Hybrid Fuzzy model uses AHP process to merge results. Pair –wise comparison of document with search engine is also performed .

The steps involved in Hybrid Fuzzy Model is described below:

Step 1) Hybrid Fuzzy model assigns Positional Values to missing documents by using H_1 heuristic as proposed by Diaz which is discussed in OWA model.

Step 2) Hence for each search engine , homogeneous list of documents positional value is obtained.

Step 3) Then AHP is applied to evaluate search engine scores. Result list of each search engine is analyzed and relationship matrix for documents is created.

Step 4) Search engine relationship matrix is used to derive document score using AHP .

Step 5) At last OWA operator is used for aggregation and normalized document score is generated.

2.6.1 Model Evaluation

Hybrid Fuzzy model performance is evaluated by De and Diaz , using datasets offered by TREC(Text retrieval conference) i.e. TREC 3, TREC 5 and TREC 9 [42]. The dataset contains 50 test queries and set of search systems. Diaz considered precision value as a metric to evaluate and compare Hybrid Fuzzy model performance with OWA model performance which is shown in Figure 6. Tests were performed by considering different quantifier values (α) i.e. 0.25 , 0.5 , 1 , 2 , 2.5 and 5. It can be clearly observed that Hybrid Fuzzy model performance is better when compared with OWA model in terms of precision.

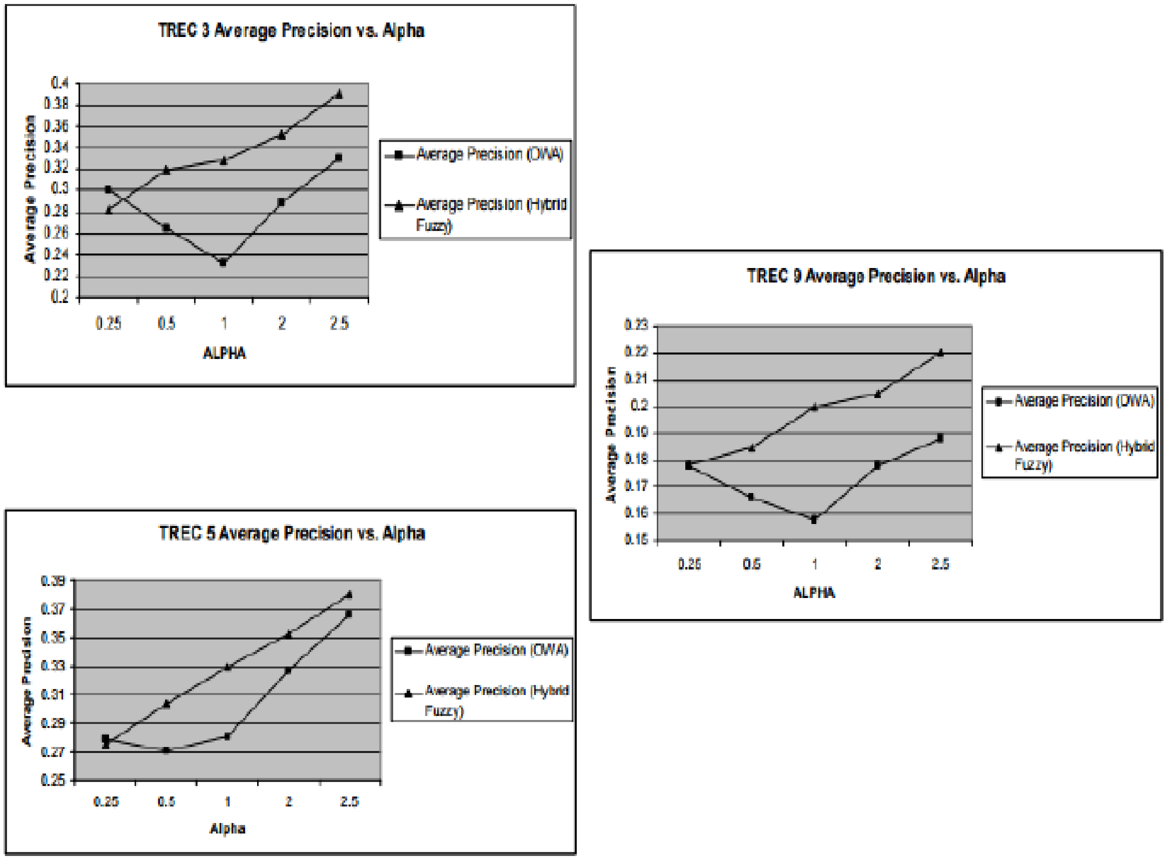


Figure 6: Hybrid Fuzzy Model Evaluation[42]

2.7 Research MSEs

Commercially available MSE such as Dogpile uses hybridized combination of parallel and serial techniques to perform metasearch[43].User query is pre-processed and dispatched to multiple search engines. Certain intelligent processing algorithm such as duplicate detection and removal, ranking etc. is applied onto returned results and document in decreasing order of preference is displayed to user. Another commercially available MSE Infospace does intelligent predictive search analysis on returned results to rank the documents[43].

2.8 Genetic Algorithm

Genetic algorithm is a heuristic search based technique that is inspired by Darwin's theory and mimics natural evolution process[50]. Genetic Algorithm (GA) is a global optimization probabilistic algorithm that simulates the process of inheritance and evolution [17]. GA is mainly used to solve problems that require expensive solutions. The flowchart of genetic Algorithm is shown in Figure 7.

GA search space contains candidate solutions to the problem known as initial population. Each candidate is represented by a string known as chromosome. Next we apply various evolutionary operators such as crossover , mutation and selection to produce next generation. Fitness function is used as an objective function for each chromosome in GA. We iterate the initial population until optimum solution is obtained or maximum number of generations have been reached[44].

The GA has been applied in various domains of real world, for example: Multi objective optimization [45], Feature selection by applying multi-objective genetic algorithms [46], Job Scheduling problems [47], Wireless Sensor Networks [48],and Cloud Computing [49] etc.

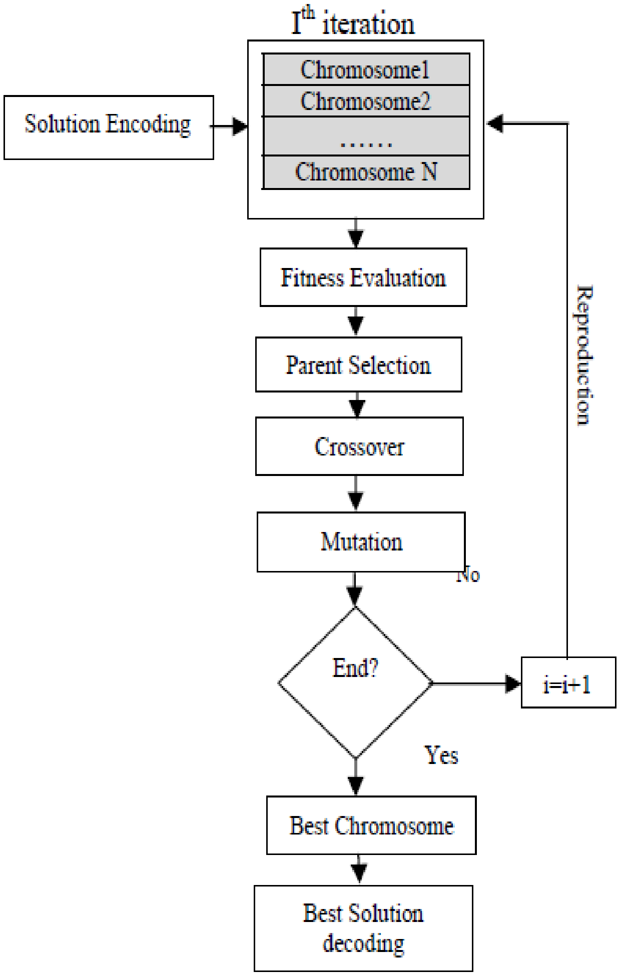


Figure 7: Flowchart of Genetic Algorithm

2.8.1 Genetic Algorithm for Optimization

Genetic Algorithm (GA) is one of the novel optimization algorithm, which is based on the concept of natural evolution and try to improve the process so that we get better results. In optimization process we apply different variations on initial idea and use the gained information to conjure up a new idea and optimize results. Genetic Algorithm is a nature inspired algorithm which uses natural selection and natural genetics method to generate optimized result.

Advantages of Genetic algorithm in field of optimization problems is listed below:

1) Genetic Algorithm uses evolution operators in their process. Evolution operators makes Genetic Algorithm more effective and efficient to perform global search. Whereas traditional algorithms uses convergent stepwise technique which compares nearby local points to perform local search.

2) Genetic Algorithm requires less mathematical computations as compared to other traditional approaches . GA have evolutionary nature which enables them to search solution in global manner without considering specific inner working of problem.

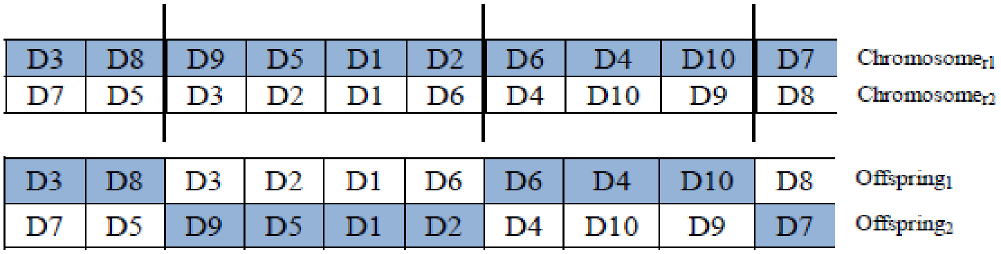
3) Genetic Algorithm brings flexibility in hybridizing domain dependent heuristics . Hence , provides efficient and effective solution to problem.

2.8.2 Result Merging using GA

The steps involved in genetic algorithm is described below:

Step 1) Generate initial population- The output of Fuzzy AHP is taken as beginning points in genetic space and then we begin to search for the best solution.

Step 2) Crossover- Crossover is a genetic operator used to exploit the potential of current population by generating offspring chromosomes. We usually select pairs of parent and apply the operator to produce children. In our proposed *MetaFusion* model , three point crossover is used. In three point crossover, three crossover points are selected and the part of the chromosome string between these three points is then swapped to generate two offspring chromosomes which is shown in Figure 8.



D3	D8	D9	D5	D1	D2	D6	D4	D10	D7	Chromosome$_{r1}$
D7	D5	D3	D2	D1	D6	D4	D10	D9	D8	Chromosome$_{r2}$
D3	D8	D3	D2	D1	D6	D6	D4	D10	D8	Offspring$_1$
D7	D5	D9	D5	D1	D2	D4	D10	D9	D7	Offspring$_2$

Figure 8: Crossover Operation

Step 3) Mutation- Mutation operator is analogous to biological mutation and is applied to maintain diversity of population. It prevents population from stagnating . In our proposed work we have used polynomial mutation[51]. The mutant vector, *m , for population p* is generated as given in Eq. (1).

$$m = \begin{cases} p + \delta_L (p - X^{(L)}), & u \leq 0.5 \\ p + \delta_R (X^{(R)} - p), & u > 0.5 \end{cases} \qquad ...(1)$$

Where

$$\delta_L = (2u)^{1/(1+\eta)} - 1 \qquad ...(2)$$

$$\delta_R = 1 - (2(1-u))^{1/(1+\eta)} \qquad ...(3)$$

Here *u* is a random number created within [0, 1] while $X^{(L)}$ and $X^{(R)}$ are allowed lower and upper bounds of documents respectively on search engines. The value of η was chosen 20 as suggested in [51].

Step 4) Selection/Reproduction - In the proposed *MetaFusion* model , OWA operator have been used as a fitness function to generate the single consolidated rank list for the user query. Fitness function is used to measure and assess the quality of an individual in current population. The selection of fitness function should be done carefully to suit the problem at hand because it is crucial for the functioning of Genetic Algorithm . " Survival of fittest" concept is used to select the best individual. For example, if jth individual in population has less fitness value than jth individual in offspring population , then, offspring individual replaces the corresponding parent individual in population.

2.9 MetaSurfer

The MetaSurfer [15] model was proposed by Tayal et al. and is based on Fuzzy AHP and Modified EOWA operator. Search engine's importance degrees are represented in terms of linguistic quantifier by using Modified EOWA . The control flow of final document ranking is shown in Figure 9.

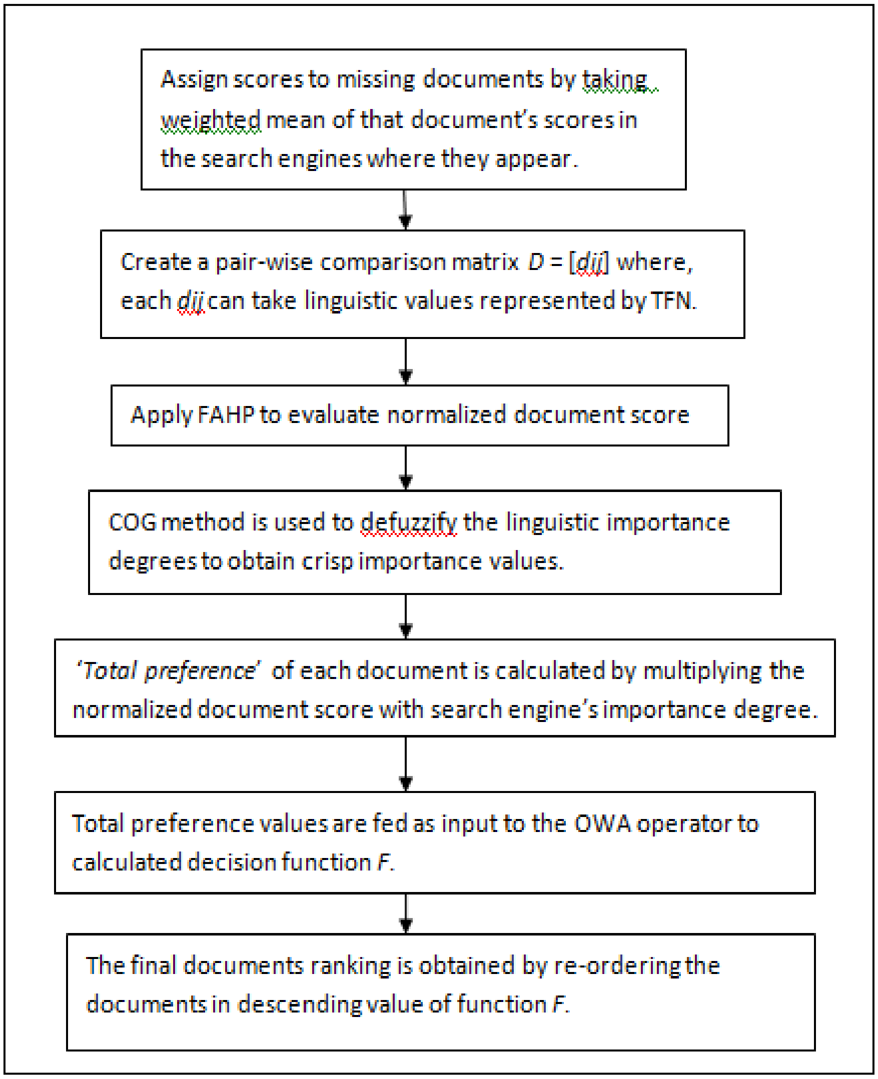

Assign scores to missing documents by taking weighted mean of that document's scores in the search engines where they appear.
Create a pair-wise comparison matrix *D* = [*dij*] where, each *dij* can take linguistic values represented by TFN.
Apply FAHP to evaluate normalized document score
COG method is used to defuzzify the linguistic importance degrees to obtain crisp importance values.
'Total preference' of each document is calculated by multiplying the normalized document score with search engine's importance degree.
Total preference values are fed as input to the OWA operator to calculated decision function *F*.
The final documents ranking is obtained by re-ordering the documents in descending value of function *F*.

Figure 9: MetaSurfer final docment ranking control flow[15]

2.9.1 Model Evaluation

Tayal et al. [15] proposed a new metric named as Weighted Precision for calculating the effectiveness of MSE. Weighted precision measures the degree of relevancy of topmost retrieved documents. The performance of MetaSurfer was compared with commercially available MSEs such as Mamma, WebCrawler and Excite in terms of precision and weighted precision over set of 14 queries. It can be clearly concluded from Table 1 that MetaSurfer has highest value of precision i.e. 2.13 and weighted precision i.e. 19.97. Note that the precision values used here are not normalized.

Table 1: MetaSurfer Evaluation[40]

Metasearch Engine	**Mean Weighted Precision**	**Mean Precision**
MetaSurfer	19.97	2.13
Mamma	13.64	1.69
WebCrawler	17.63	1.88
Excite	19.00	1.97

2.10 MetaXplorer

Very recent model of MSE ,MetaXplorer [16], was proposed by Daya Gupta and Neha Dimri to allow for performing metasearch on user query. The steps involved in ranking computation is described below:

1. User posted query is refined and dispatched to several underlying search engines.
2. In-OWA operator was used to assign importance degree to search engines.
3. Missing document’s are assigned weight by taking weighted mean of that document in those search engines where they appear.
4. FAHP is applied to perform pair-wise comparison of documents and hence document scoring is done.

5. The overall document preference is obtained by multiplying document score with search engine's importance degree.

2.11 Comparative Analysis of the Surveyed Models

Several existing models of meta-search are studied and analyzed and their comparison is shown in Table 2.

Table 2: Summarization Of MetaSearch Models

MetaSearch model	**Year of establishment**	**Underlying techniques used**	**Major Advantages**	**Main Shortcomings**
MetaCrawler	1995	Confidence factor evaluation using a voting scheme	Simple method, Query formulation specific to search services, duplicate removal	Search engines considered to be equally important
Borda-Fuse Model	2001	Borda Count voting algorithm	Straightforward technique, Allows search engines to vote	Search engines considered to be equally important

Weighted Borda-Fuse	2001	Borda points along with search engine weights	Considers heterogeneous search environment	Missing documents are assigned lesser Borda points
OWA model	2005	OWA based Aggregation	Proposed two heuristics for missing documents	Does not consider inter-document relationships or search engines' similarity
Hybrid Fuzzy Model	2009	AHP and OWA operator	Performs pair-wise comparison of documents as well as search engines	OWA sometimes performs worse than T-norm OWA operator
MetaSurfer	2014	FAHP and modified EOWA	Slightly different heuristic for missing documents, linguistic comparisons are made, linguistic importance degrees	Does not consider the dynamic nature of the Web, Documents are ranked just on the basis of search engine preferences

MetaXplorer	2015	FAHP and In-OWA	Intelligent OWA operator is used to solve MCDM problem	Binary classification is used to measure documents relevance

CHAPTER 3: PROPOSED METASEARCH MODEL

This chapter presents our proposed model for meta-search, *MetaFusion*. The proposed approach consists of two phases i.e. Training Phase and Query Execution Phase. The proposed algorithm uses Fuzzy AHP along with Genetic Algorithm for result merging. URL Analysis is also performed to analyze each document's URL. Hence ,it makes *MetaFusion* to be intelligent.The proposed model is dynamic and free from expert's biased opinion. This chapter also discusses the advantages of our proposed model, *MetaFusion*, over previously existing models. The details regarding implementation and performance evaluation of *MetaFusion* will be discussed in chapters 4 and 5 respectively.

3.1 Proposed Model for Meta-Search: MetaFusion

The proposed *MetaFusion* model basically have two phases i.e. Training phase and Query Execution phase. In training phase, we give training examples as input to training algorithm and we get search engine's importance degree weight as output. After the completion of training phase query execution phase happens. In query execution phase fuzzy AHP and Genetic algorithm is applied to form consolidated rank list. The working of two phases is discussed below.

3.1.1 Training Phase

The training phase consist of training algorithm to which training examples are fed as input and cumulative importance degrees of underlying search engines are generated as output which is shown in Figure 10.

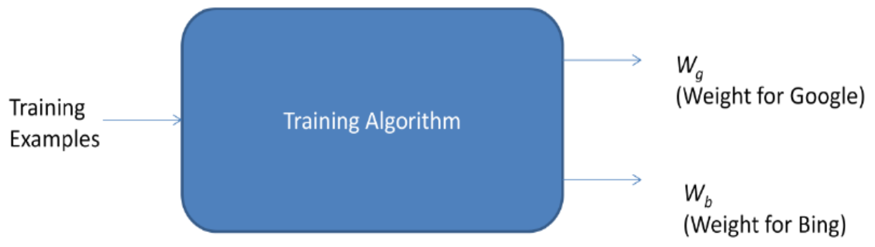


Figure 10: Training Phase

The training example basically consist of different queries , ranking of document's according to each search engine and optimal ranking of documents.

For developing adaptable Meta-Search Engine, we need a model which adapts to environmental changes. Search engine's performance varies with time due to the updation of ranking and indexing algorithms. To reflect these performance changes we need an adaptable meta-search model which assigns importance degree weights to search engines in unbiased manner.

Consider an example where Yahoo and Bing are used as underlying search engines where user query is forwarded. In beginning Yahoo performed better than Bing and therefore , higher importance weight is assigned to Yahoo. But after some time, Bing may update its searching algorithm , thereby, giving better results than Yahoo. Now at this point Bing should be assigned higher importance degree than Yahoo. Therefore , we need a model which automatically updates search engines importance degree weight from time to time.

In our proposed model search engines are considered as criteria and documents as alternatives. Google and Bing are the two underlying search engines where user query is dispatched and results are fetched. Thus , the cumulative importance degrees of search engines i.e. Google(W_g) and Bing (W_b) are computed in training phase . Hence this makes our model heterogeneous in nature and adaptable because it can adapt to changing environment. We can run training algorithm periodically or as per user feedback to make our model flexible and updated.

3.1.2 Query Execution Phase

Query execution phase is invocated when user submits a query through MetaFusion interface. Query Execution Phase consist of various modules : preprocessing module, URL Analysis module, Google query computation module, and Bing Query computation module. MetaFusion applies Fuzzy AHP and Genetic algorithm to form a consolidated rank list. The whole working of Query Execution Phase is shown in Figure 11.

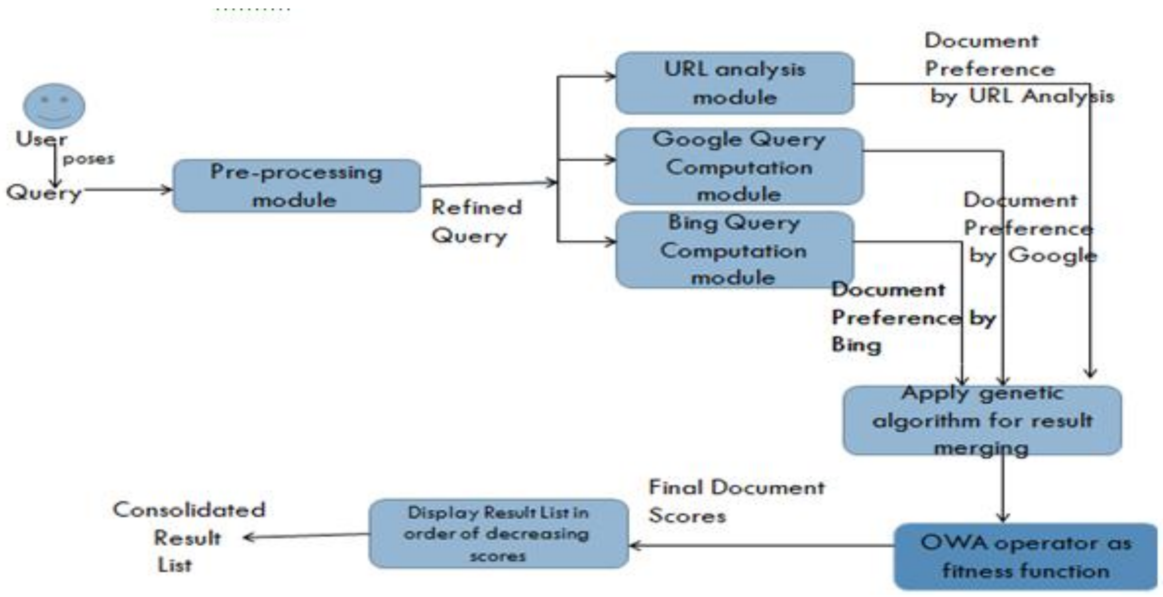

User
poses
Query
Pre-processing module
Refined Query
URL analysis module
Google Query Computation module
Bing Query Computation module
Document Preference by URL Analysis
Document Preference by Google
Document Preference by Bing
Apply genetic algorithm for result merging
OWA operator as fitness function
Final Document Scores
Display Result List in order of decreasing scores
Consolidated Result List

Figure 11: Working of Query Execution Phase

The steps involved in Query Execution Phase is described below:

i. *Preprocessing Module*:

In this module query preprocessing happens by removing stop words and redundant terms . Stop words are common words like a, an, the, or, for etc. which are filtered out from search query because they slow down the process of result extraction without improving their accuracy. Thus, generated refined query is dispatched to underlying search engines i.e. Google and Bing.

ii. *URL Analysis Module*:

URL Analysis module analyses and inspects the URL of each documents to determine their relevancy. “Document preference” of returned documents is computed by assigning higher weight values to more relevant documents . The contents of research paper or journal is more relevant than contents belonging to textbook. Similarly, textbook chapter provides more relevant content than ordinary dictionary website explaining the meaning of submitted query. Hence , URL Analysis makes our proposed MSE to be intelligent because we are not simply aggregating the returned results.

The “Document Preference” weight *DW* for each document in URL Analysis is assigned in the following manner:

a. *DW* is assigned to 0.4 if it belongs to abstract or full text of research journal or conference paper.

b. *DW* is assigned to 0.3 if it belongs to journal or conference homepage.

c. *DW* is assigned to 0.2 if it belongs to database i.e Wikipedia or book.

d. *DW* is assigned to 0.1 if it does not belongs to above mentioned category i.e. company web pages, dictionaries etc.

iii. *Google Query Computation Module*:

The working of Google Query Computation module is shown in Figure 12.

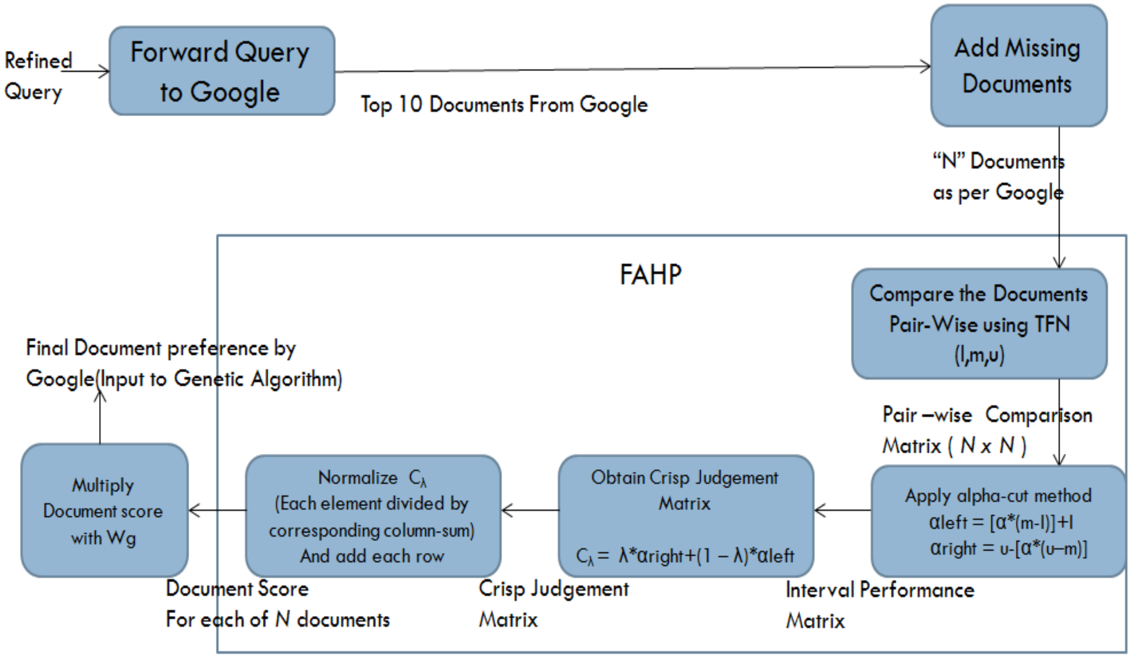

Refined Query
Forward Query to Google
Top 10 Documents From Google
Add Missing Documents
"N" Documents as per Google
FAHP
Compare the Documents Pair-Wise using TFN (l,m,u)
Pair –wise Comparison Matrix (N x N)
Apply alpha-cut method
αleft = [α*(m-l)]+l
αright = u-[α*(u–m)]
Interval Performance Matrix
Obtain Crisp Judgement Matrix
Cλ = λ*αright+(1 – λ)*αleft
Crisp Judgement Matrix
Normalize Cλ
(Each element divided by corresponding column-sum)
And add each row
Document Score For each of N documents
Multiply Document score with Wg
Final Document preference by Google(Input to Genetic Algorithm)

Figure 12: Google Query computation

The steps involved are discussed below:

Step 1: The generated refined query is dispatched to Google and then top ten result list documents are fetched.

Step 2: The documents that are missing in Google but present in Bing are added to Google's result list to generate a list of 'N' documents according to Google, where N denotes total number of unique documents which are considered by taking Google and Bing result together. We add missing documents in Google result list by calculating their weighted average of positioning in each search engine's list.

Step 3: Then document score is evaluated by applying Fuzzy AHP Algorithm[30].The linguistic variables are used to form pair –wise comparison matrix of size N×N is shown below. Then, linguistic variables are represented by Triangular Fuzzy Number(TFN) which is denoted by (l,m,u), in which l represents left , m represents middle and u represents right component of TFN.

Least Important (LTI)	(1,1,3)
Less Important (LSI)	(1,3,5)
Equally Important (EI)	(3,5,7)
More Important (MEI)	(5,7,9)
Most Important (MSI)	(7,9,9)

Step 4: Apply alpha-cut method to form interval performance matrix [αleft, αright] and is computed as follows:

$$\alpha\text{left} = [\alpha * (m-l)] + l$$

$$\alpha\text{right} = u - [\alpha * (u - m)]$$

where α is confidence factor which ε between[0,1].

Step 5: Obtain Crisp Judgement Matrix ,C_λ , by using following equation:

$$C_\lambda = \lambda * \alpha\text{right} + (1 - \lambda) * \alpha\text{left}$$

where λ is optimism index of decision maker and λ ε[0,1].

Step 6: Then, Normalize C_λ by dividing each element by corresponding column's-sum. After that add each row to get document score , *di*.

Step 7: Obtain final preference of document by multiplying document score *di* with Google importance degree weight(*Wg*) as shown below.

$$DPi = Wg * di$$

iv. *Bing Query Computation Module*:

The steps involved in Bing query computation module is similar to the Google Query computation module. In this Fuzzy AHP is applied to documents returned by Bing search engine. The working of this module is shown in Figure 13.

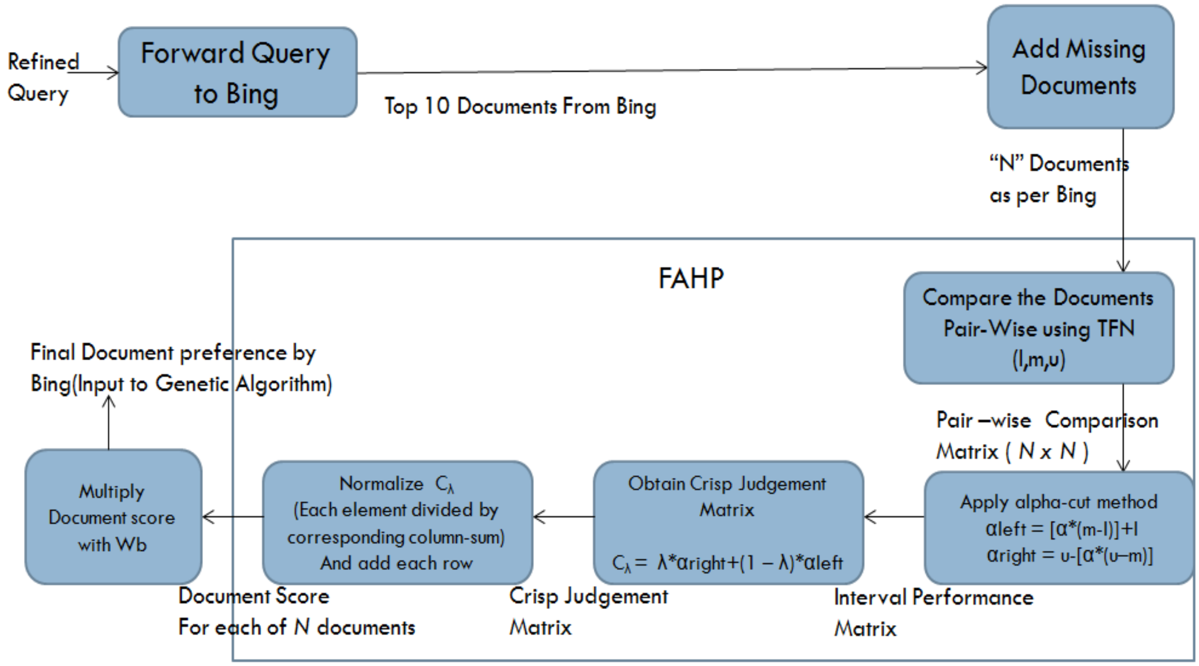

Refined Query
Forward Query to Bing
Top 10 Documents From Bing
Add Missing Documents
"N" Documents as per Bing
FAHP
Compare the Documents Pair-Wise using TFN (l,m,u)
Pair –wise Comparison Matrix (N x N)
Apply alpha-cut method
αleft = [α*(m-l)]+l
αright = u-[α*(u–m)]
Interval Performance Matrix
Obtain Crisp Judgement Matrix
Cλ = λ*αright+(1 – λ)*αleft
Crisp Judgement Matrix
Normalize Cλ
(Each element divided by corresponding column-sum)
And add each row
Document Score For each of N documents
Multiply Document score with Wb
Final Document preference by Bing(Input to Genetic Algorithm)

Figure 13: Bing Query Computation

v. *Genetic Algorithm module* :

The Genetic Algorithm (GA) module merges the result and form the single consolidated rank list of documents based on the fitness value of document. Genetic Algorithm uses evolutionary operators i.e. crossover, mutation and selection to perform Global search. Genetic Algorithm is applied for obtaining optimized results.

The steps involved in Genetic Algorithm is shown below:

Step 1: Calculate OWA operator weight using

$$W_i = \left(\frac{i}{n}\right)^{\propto} - \left(\frac{i-1}{n}\right)^{\propto}$$

where n is number of criteria and α ε[0,1].

Step 2: Now compute the *Ordered Weighted Average* and average for same using

$$OWA_{i,j} = \sum_{j=1}^{N} W_i \times dscore_{i,j}$$

where *dscore* is document score obtained by FAHP.

$$Avg\left(OWA_{i,j}\right) = \frac{\sum_{k=1}^{X^U} OWA_k}{X^U}$$

where X^U is maximum number of documents a search engine can return.

Step 3: Fitness of an individual document is obtained as follows[57]:

$$fitness(i) = \begin{cases} \beta, CP > fitness(i) \| fitness(i) == \max Fit \\ \left| \dfrac{Avg(OWA)_i - CP}{\max Fit - CP} \right|, otherwise \end{cases}$$

where β ε[0,1] is a real random number, CP denotes cut point as *0.5×maxFit* and *maxFit* is the maximum value of *Avg OWA* .

Step 4: Apply crossover operation and generate offspring chromosomes. In proposed work three point crossover is used.

Step 5: Apply polynomial mutation operation over generated offspring vector to bring diversity in population

Step 6: Evaluate newly generated chromosomes and select best amongst them based on basis of their fitness value i.e. “survival of fittest”.

CHAPTER 4 : IMPLEMENTATION

This chapter provides the details regarding to the implementation of the proposed *MetaFusion* model. Firstly brief description of proposed work's implementation platform and names of various used jar files is described. Then implementation details of training phase and query execution phase is described.

4.1 Brief Description of Implementation

The proposed meta-search engine model, MetaFusion , is implemented by using Netbeans IDE 6.9.1 , JAVA EE 6 and MATLAB R2007b platform. The major steps involved in implementation is described below:

- Design the interface of proposed model, MetaFusion
- User poses query into MetaFusion interface.
- MetaFusion preprocesses the query and forwards it to underlying search engines i.e. Bing and Google.
- Train the proposed MetaFusion model to consider search engine's importance weight.
- Perform URL analysis on returned results.
- Add missing documents in the Google and Bing's result list.
- Apply Fuzzy AHP to reflect human thinking.
- Apply Genetic algorithm for multi-optimization.
- Display the documents in decreasing value of Fitness Function.

The JAR (JAVA Archive) files used in our proposed model, MetaFusion , are described below:

- *Google API Services Custom Search 1.20.0* [52] is used to forward user query to Google and fetch results.
- Netbeans IDE 6.9.1 interacts with MATLAB using *MatlabControl 4.1.0* [53] jar file.
- Google Custom Search API dependencies are resolved by *HttpClient 4.1* [54] jar file.
- Google Custom Search API dependencies are resolved by *HttpCore 4.1* [55] jar file.
- Google Custom Search API dependencies are resolved by *HttpMime 4.1* [56] jar file.
- *Azure Bing Search Java 0.12.0* [57] allows to forward user query to Bing and fetch results.
- Bing Search API dependencies are resolved by *Org Apache Commons logging* [58] jar file.
- Bing Search API dependencies are resolved by *Org Apache Commons codec* [59] jar file.
- Bing Search API dependencies are resolved by *Org Apache Commons net 3.3* [60] jar file.

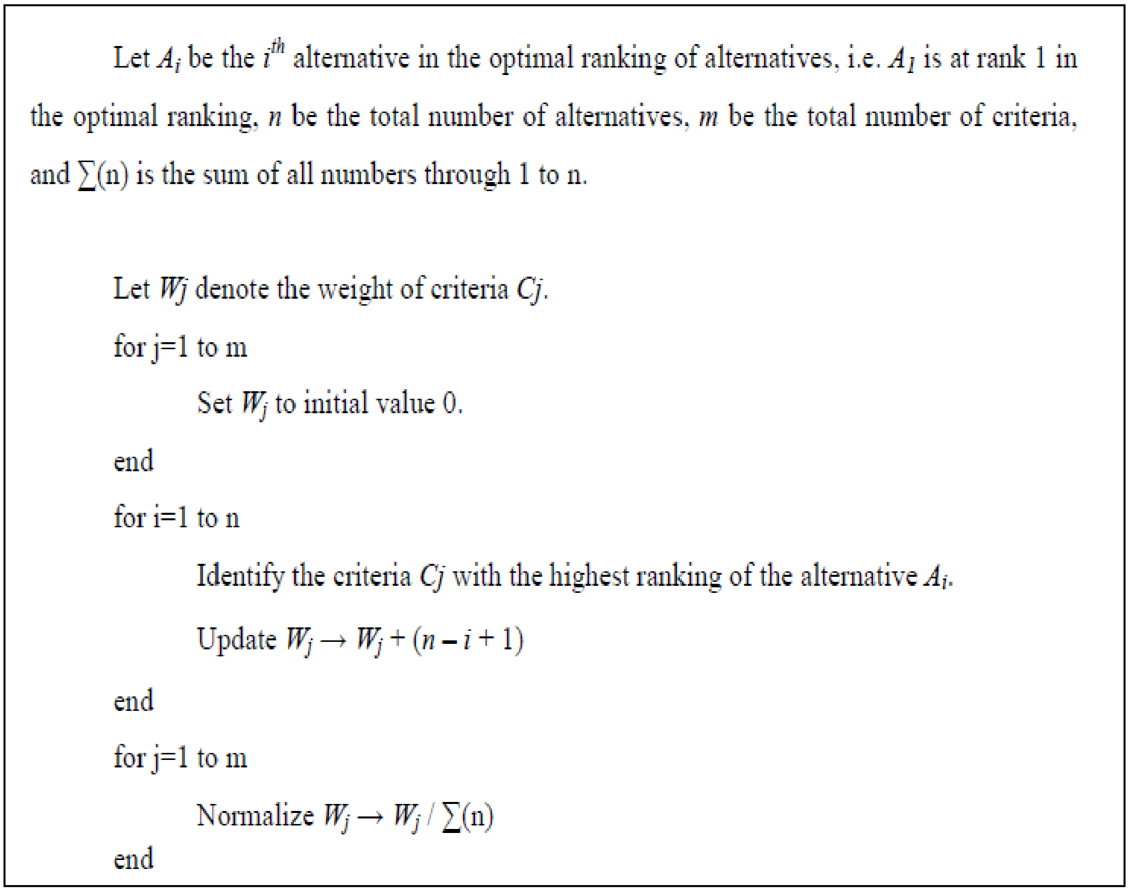
4.2 Training phase implementation details

The training phase of proposed model, MetaFusion , consists of training examples. Each training example consist of document ranking according to Bing, Google and optimal ranking. In our proposed model we have considered ten example queries to train our MetaFusion which are shown below:

- Ontology
- Deep Learning
- Information Retrieval
- Remote Sensing
- Genetic Algorithm
- Data Mining Techniques

- Cryptography
- Job Scheduling
- Prediction Neural Network.
- Biogeography Based Optimization

The importance degree to search engine's are assigned using learning algorithm. In our proposed model, experts don't assign weights to search engines thereby making our process biased free. The working of learning algorithm is shown in Figure 14.

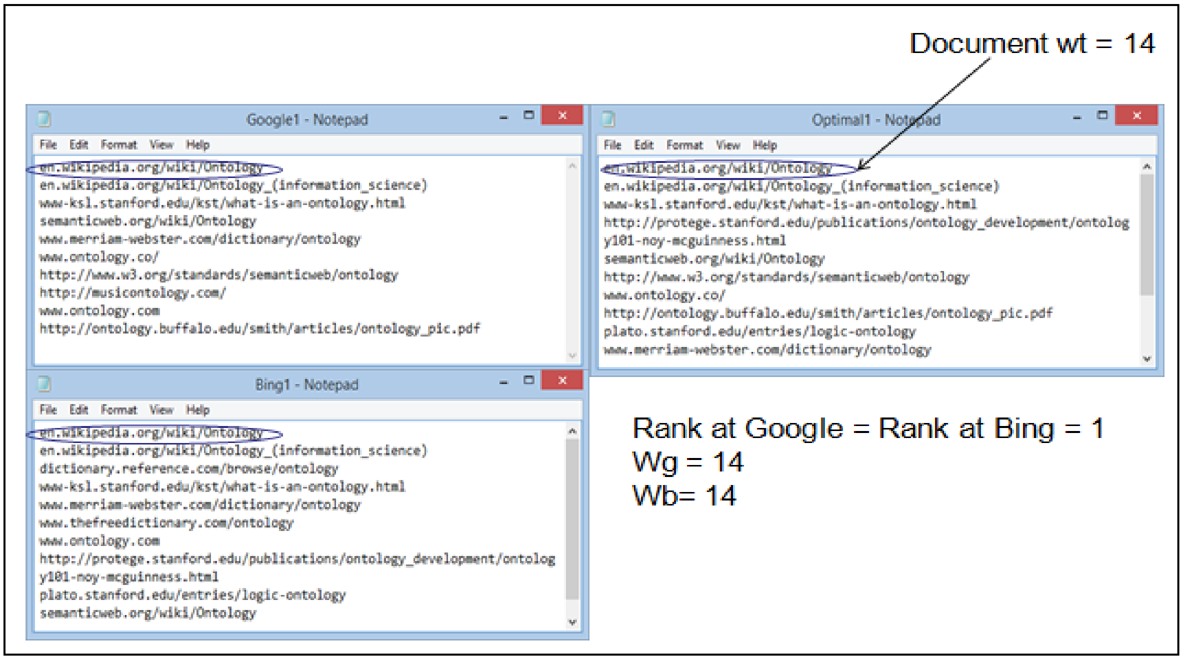


Let A_i be the i^{th} alternative in the optimal ranking of alternatives, i.e. A_1 is at rank 1 in the optimal ranking, n be the total number of alternatives, m be the total number of criteria, and $\sum(n)$ is the sum of all numbers through 1 to n.

```
Let $Wj$ denote the weight of criteria $Cj$.
for j=1 to m
    Set $W_j$ to initial value 0.
end
for i=1 to n
    Identify the criteria $Cj$ with the highest ranking of the alternative $A_i$.
    Update $W_j \rightarrow W_j + (n - i + 1)$
end
for j=1 to m
    Normalize $W_j \rightarrow W_j / \sum(n)$
end
```

Figure 14: Working of Learning Algorithm

The working of training algorithm is shown with the help of example in Figure 15 , 16 ,17 respectively. Consider the query as ‘ontology’.


Google1 - Notepad
File Edit Format View Help
Bing1 - Notepad
File Edit Format View Help
Optimal1 - Notepad
File Edit Format View Help

Document wt = 14

Google1 - Notepad

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
www-ksl.stanford.edu/kst/what-is-an-ontology.html
semanticweb.org/wiki/Ontology
www.merriam-webster.com/dictionary/ontology
www.ontology.co/
http://www.w3.org/standards/semanticweb/ontology
http://musicontology.com/
www.ontology.com
http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
```

Bing1 - Notepad

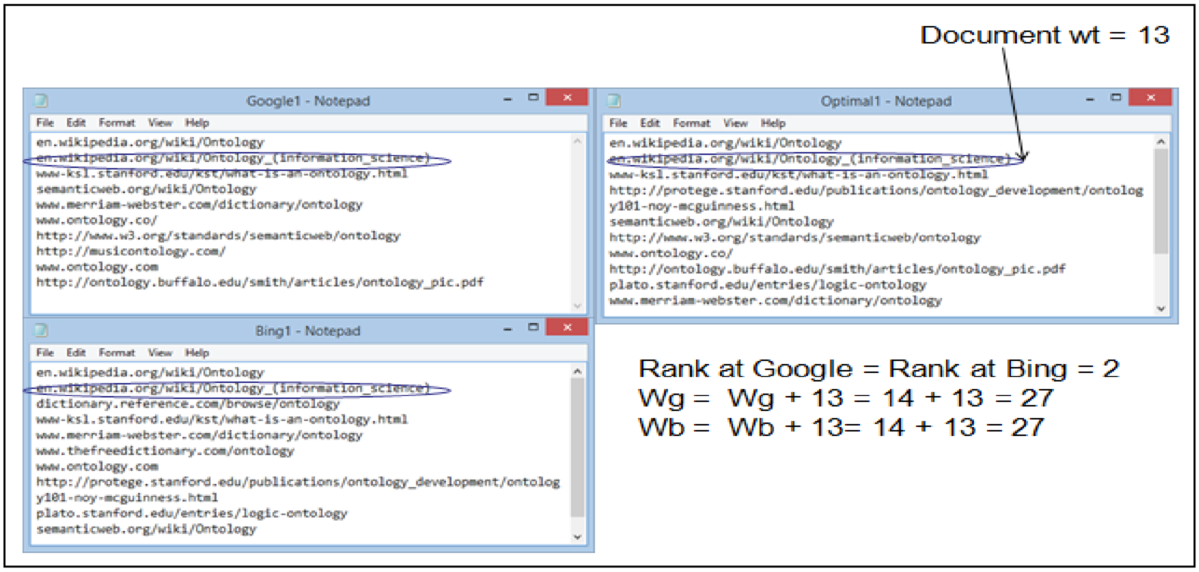
```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
dictionary.reference.com/browse/ontology
www-ksl.stanford.edu/kst/what-is-an-ontology.html
www.merriam-webster.com/dictionary/ontology
www.thefreedictionary.com/ontology
www.ontology.com
http://protege.stanford.edu/publications/ontology_development/ontolog
y101-noy-mcguinness.html
plato.stanford.edu/entries/logic-ontology
semanticweb.org/wiki/Ontology
```

Optimal1 - Notepad

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
www-ksl.stanford.edu/kst/what-is-an-ontology.html
http://protege.stanford.edu/publications/ontology_development/ontolog
y101-noy-mcguinness.html
semanticweb.org/wiki/Ontology
http://www.w3.org/standards/semanticweb/ontology
www.ontology.co/
http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
plato.stanford.edu/entries/logic-ontology
www.merriam-webster.com/dictionary/ontology
```

Rank at Google = Rank at Bing = 1
Wg = 14
Wb= 14

Figure 15: Working of Training Phase- initial step


Google1 - Notepad
File Edit Format View Help
Bing1 - Notepad
File Edit Format View Help
Optimal1 - Notepad
File Edit Format View Help

Document wt = 13

Google1 - Notepad

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
www-ksl.stanford.edu/kst/what-is-an-ontology.html
semanticweb.org/wiki/Ontology
www.merriam-webster.com/dictionary/ontology
www.ontology.co/
http://www.w3.org/standards/semanticweb/ontology
http://musicontology.com/
www.ontology.com
http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
```

Bing1 - Notepad

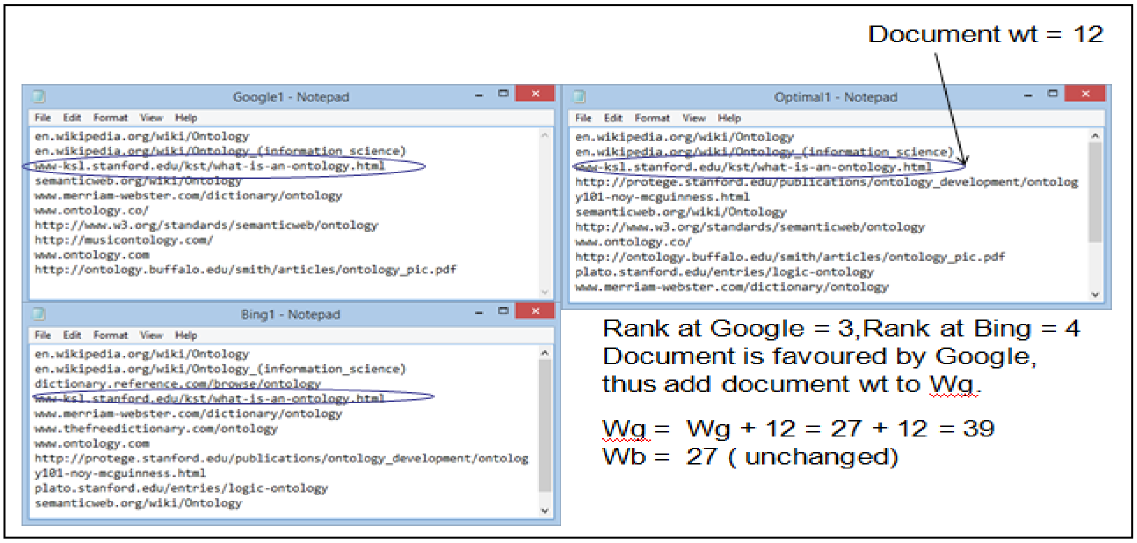
```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
dictionary.reference.com/browse/ontology
www-ksl.stanford.edu/kst/what-is-an-ontology.html
www.merriam-webster.com/dictionary/ontology
www.thefreedictionary.com/ontology
www.ontology.com
http://protege.stanford.edu/publications/ontology_development/ontolog
y101-noy-mcguinness.html
plato.stanford.edu/entries/logic-ontology
semanticweb.org/wiki/Ontology
```

Optimal1 - Notepad

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
www-ksl.stanford.edu/kst/what-is-an-ontology.html
http://protege.stanford.edu/publications/ontology_development/ontolog
y101-noy-mcguinness.html
semanticweb.org/wiki/Ontology
http://www.w3.org/standards/semanticweb/ontology
www.ontology.co/
http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
plato.stanford.edu/entries/logic-ontology
www.merriam-webster.com/dictionary/ontology
```

Rank at Google = Rank at Bing = 2
Wg = Wg + 13 = 14 + 13 = 27
Wb = Wb + 13= 14 + 13 = 27

Figure 16: Working of Training Phase- second step


Google1 - Notepad
File Edit Format View Help
Optimal1 - Notepad
File Edit Format View Help
Bing1 - Notepad
File Edit Format View Help

Document wt = 12

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
www-ksl.stanford.edu/kst/what-is-an-ontology.html
semanticweb.org/wiki/Ontology
www.merriam-webster.com/dictionary/ontology
www.ontology.co/
http://www.w3.org/standards/semanticweb/ontology
http://musicontology.com/
www.ontology.com
http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
```

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
www-ksl.stanford.edu/kst/what-is-an-ontology.html
http://protege.stanford.edu/publications/ontology_development/ontolog
y101-noy-mcguinness.html
semanticweb.org/wiki/Ontology
http://www.w3.org/standards/semanticweb/ontology
www.ontology.co/
http://ontology.buffalo.edu/smith/articles/ontology_pic.pdf
plato.stanford.edu/entries/logic-ontology
www.merriam-webster.com/dictionary/ontology
```

```
en.wikipedia.org/wiki/Ontology
en.wikipedia.org/wiki/Ontology_(information_science)
dictionary.reference.com/browse/ontology
www-ksl.stanford.edu/kst/what-is-an-ontology.html
www.merriam-webster.com/dictionary/ontology
www.thefreedictionary.com/ontology
www.ontology.com
http://protege.stanford.edu/publications/ontology_development/ontolog
y101-noy-mcguinness.html
plato.stanford.edu/entries/logic-ontology
semanticweb.org/wiki/Ontology
```

Rank at Google = 3,Rank at Bing = 4
Document is favoured by Google,
thus add document wt to Wg.

Wg = Wg + 12 = 27 + 12 = 39
Wb = 27 (unchanged)

Figure 17: Working of training phase- third step

Repeat this process all the 14 documents in Optimal Ranking to obtain Wg and Wb as:

$$Wg = 82$$

$$Wb = 55$$

Normalize Weights:

$$Wg = Wg / \sum(14) = 82/105 = 0.78$$

$$Wb = Wb / \sum(14) = 55/105 = 0.52$$

4.3 Query execution Phase Implementation details

We have designed a user friendly interface for our proposed model, MetaFusion , which is shown in Figure 18 . The interface allows user to submit his query. When user clicks ‘Search’ button, MetaFusion preprocesses the query and dispatches it to underlying search engines i.e. Google and Bing.

In our proposed work we have fetched results from Google using Google Custom Search API[47] and Bing results are fetched using ,Bing Search API [48], published by Microsoft.

Different payment plans are available for both the services based on the number of search queries sent. Google Custom Search API is free of charge allowing 100 search queries per day and Bing Search API is freely available for the limit of 5000 transactions/month .

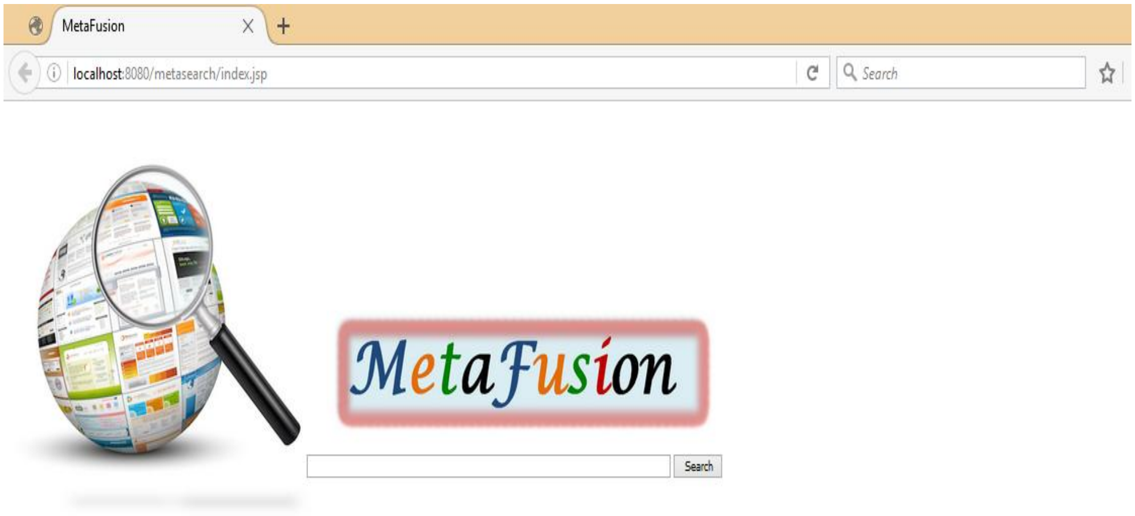


Figure 18: Proposed model “MetaFusion” Interface

CHAPTER 5: EVALUATION AND RESULTS

This chapter describes the performance evaluation of proposed meta-search model, MetaFusion. Also the results of MetaFusion is compared with results of research MSEs.

5.1 Performance Evaluation

The performance of proposed model *MetaFusion* is calculated in terms of precision . Precision is basically defined as ratio of retrieved relevant documents out of total number of documents retrieved , which is shown below:

$$Precision = \frac{The\ number\ of\ retrieved\ relevant\ documents}{The\ number\ of\ retrieved\ documents}$$

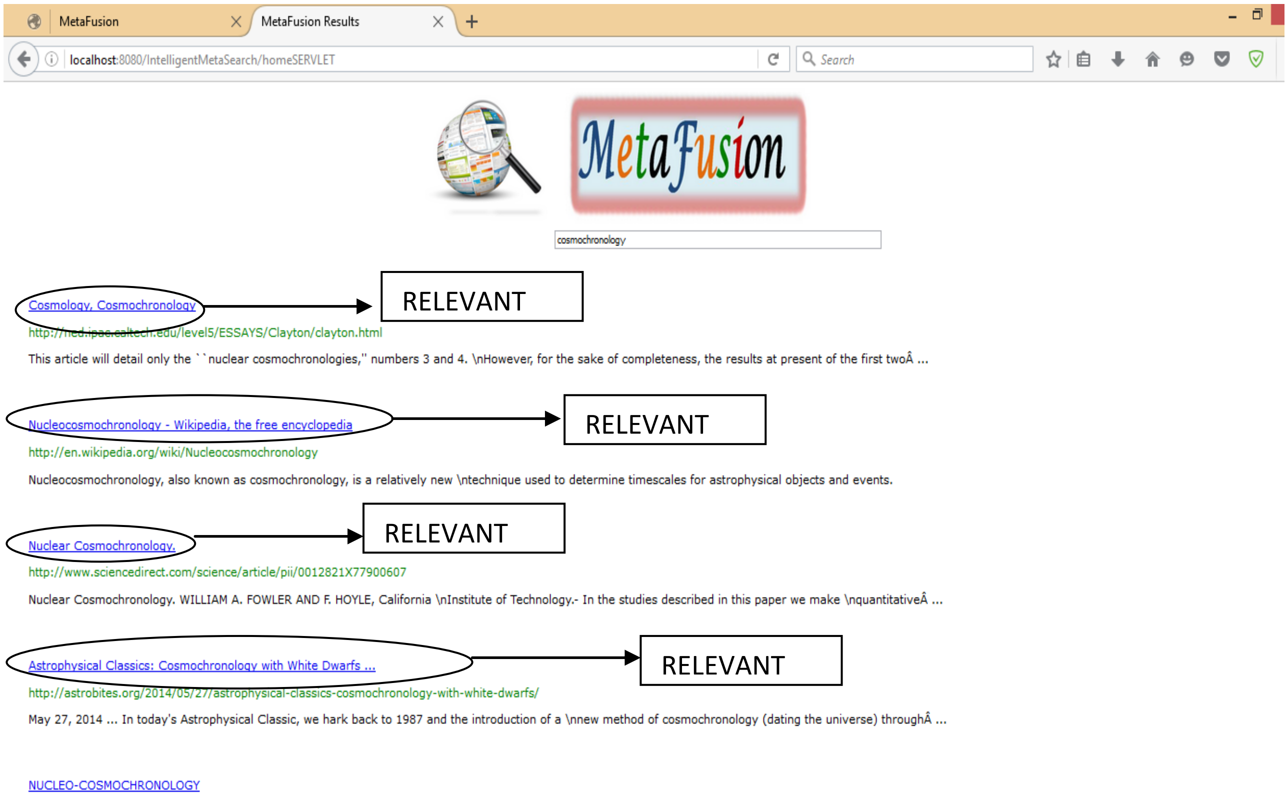
We have compared the performance of our proposed model with existing research MSEs by considering over 100 test queries from different domains of real world. Some of these sample queries are listed in Table 3.

Table 3: Some Sample queries

Machine Learning	Cosmochronology	Text Mining	Image Processing
Neurobiology	Cognitive Science	Multi Agent System	Photosynthesis
Expert System	Branch Prediction	Human genetics	Robotic Fusion
Optical technology	Image forensics	Semantic Analysis	Partical Swarm Intelligence
Branch Prediction	Pattern Recognition	Cryptography	Ontology

5.2 Results

The results obtained by MetaFusion , Dogpile, InfoSpace and PolyMeta on the query "Cosmochronology" is shown in Figure 19, 20 , 21 and 22 respectively.



Cosmology, Cosmochronology

RELEVANT

http://ned.ipac.caltech.edu/level5/ESSAYS/Clayton/clayton.html

This article will detail only the ``nuclear cosmochronologies,'' numbers 3 and 4. \nHowever, for the sake of completeness, the results at present of the first twoÂ ...

Nucleocosmochronology - Wikipedia, the free encyclopedia

RELEVANT

http://en.wikipedia.org/wiki/Nucleocosmochronology

Nucleocosmochronology, also known as cosmochronology, is a relatively new \ntechnique used to determine timescales for astrophysical objects and events.

Nuclear Cosmochronology.

RELEVANT

http://www.sciencedirect.com/science/article/pii/0012821X77900607

Nuclear Cosmochronology. WILLIAM A. FOWLER AND F. HOYLE, California \nInstitute of Technology.- In the studies described in this paper we make \nquantitativeÂ ...

Astrophysical Classics: Cosmochronology with White Dwarfs ...

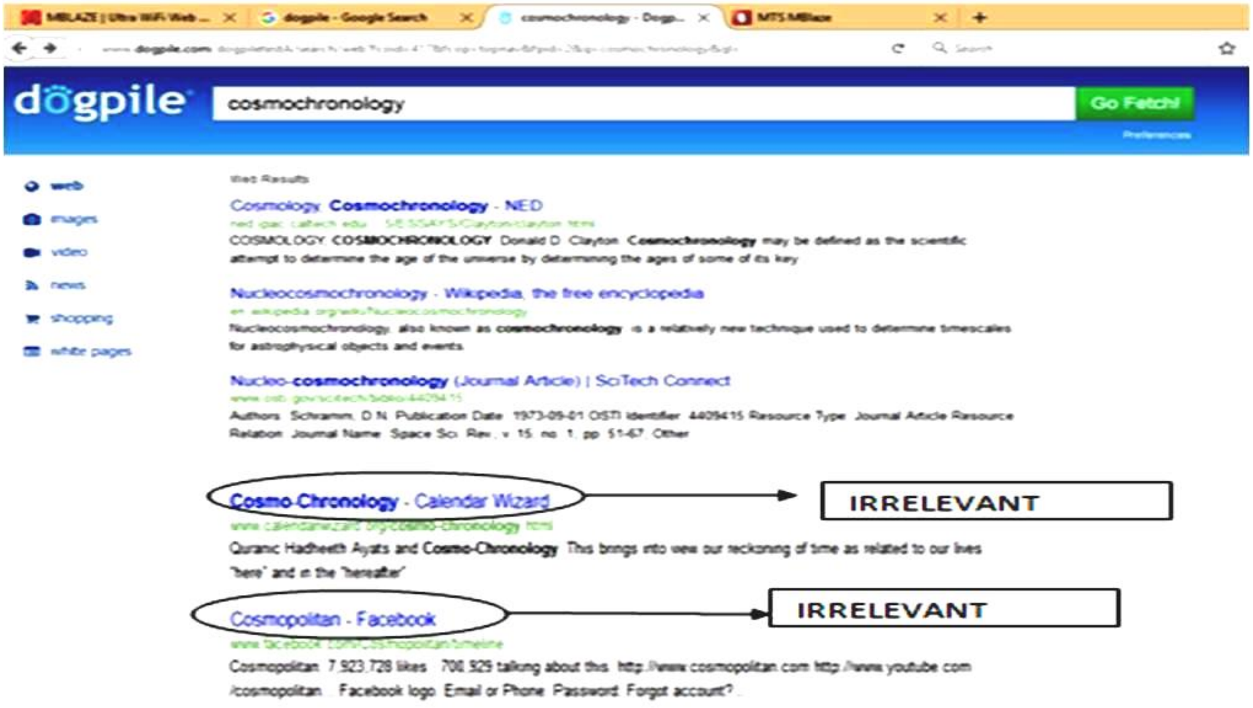
RELEVANT

http://astrobites.org/2014/05/27/astrophysical-classics-cosmochronology-with-white-dwarfs/

May 27, 2014 ... In today's Astrophysical Classic, we hark back to 1987 and the introduction of a \nnew method of cosmochronology (dating the universe) throughÂ ...

NUCLEO-COSMOCHRONOLOGY

Figure 19: Results of proposed " MetaFusion" Model



cosmochronology

Web Results

Cosmology, Cosmochronology - NED
ned.ipac.caltech.edu/.../5/ESSAYS/Clayton/clayton.html
COSMOLOGY, COSMOCHRONOLOGY. Donald D. Clayton. Cosmochronology may be defined as the scientific attempt to determine the age of the universe by determining the ages of some of its key

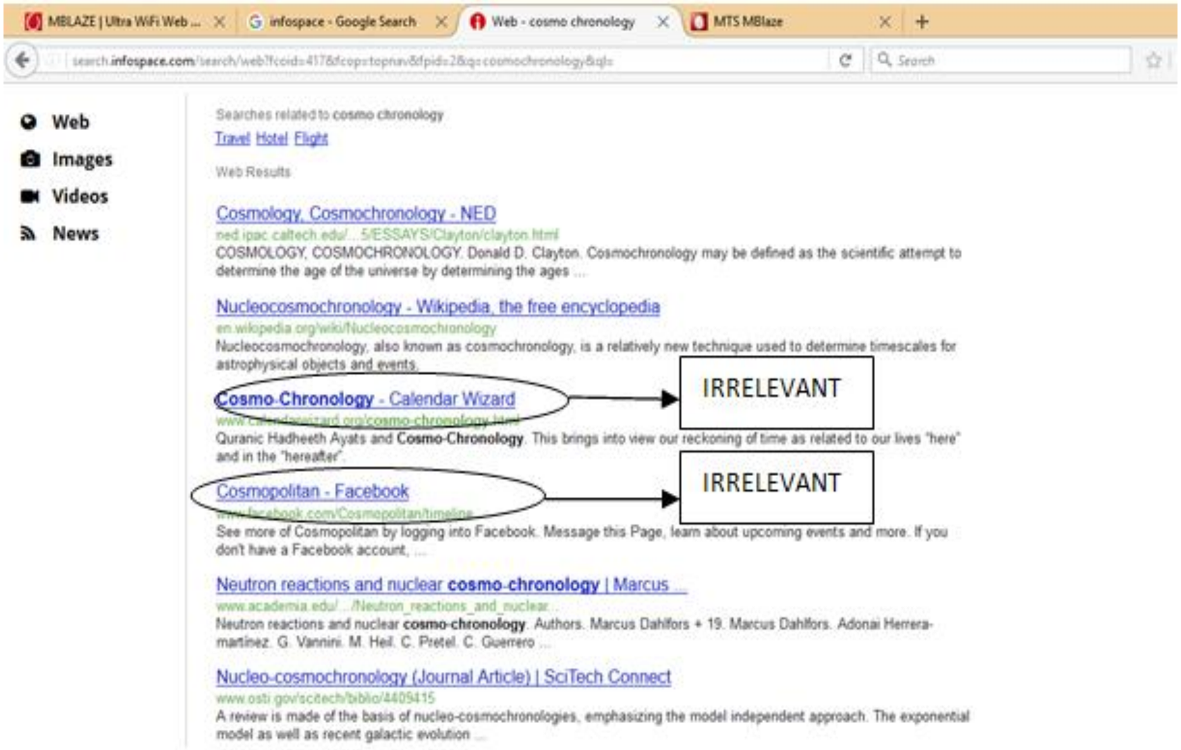
Nucleocosmochronology - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Nucleocosmochronology
Nucleocosmochronology, also known as cosmochronology, is a relatively new technique used to determine timescales for astrophysical objects and events.

Nucleo-cosmochronology (Journal Article) | SciTech Connect
www.osti.gov/scitech/biblio/4409415
Authors: Schramm, D.N. Publication Date: 1973-09-01 OSTI Identifier: 4409415 Resource Type: Journal Article Resource Relation: Journal Name: Space Sci. Rev., v. 15, no. 1, pp. 51-67. Other

Cosmo-Chronology - Calendar Wizard
www.calendarwizard.org/cosmo-chronology.html
Quranic Hadheeth Ayats and Cosmo-Chronology. This brings into view our reckoning of time as related to our lives "here" and in the "hereafter"

Cosmopolitan - Facebook
www.facebook.com/Cosmopolitan/timeline
Cosmopolitan. 7,923,728 likes · 708,929 talking about this. http://www.cosmopolitan.com http://www.youtube.com /cosmopolitan ... Facebook logo. Email or Phone. Password. Forgot account? ...

Figure 20: Dogpile Results


MBLAZE | Ultra WiFi Web ...
infospace - Google Search
Web - cosmo chronology
MTS MBlaze
search.infospace.com/search/web?fcoid=417&fcop=topnav&fpid=2&q=cosmochronology&ql=
Search
Web
Images
Videos
News
IRRELEVANT
IRRELEVANT

Searches related to cosmo chronology

Travel Hotel Flight

Web Results

Cosmology, Cosmochronology - NED
ned.ipac.caltech.edu/.../5/ESSAYS/Clayton/clayton.html
COSMOLOGY, COSMOCHRONOLOGY. Donald D. Clayton. Cosmochronology may be defined as the scientific attempt to determine the age of the universe by determining the ages ...

Nucleocosmochronology - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Nucleocosmochronology
Nucleocosmochronology, also known as cosmochronology, is a relatively new technique used to determine timescales for astrophysical objects and events.

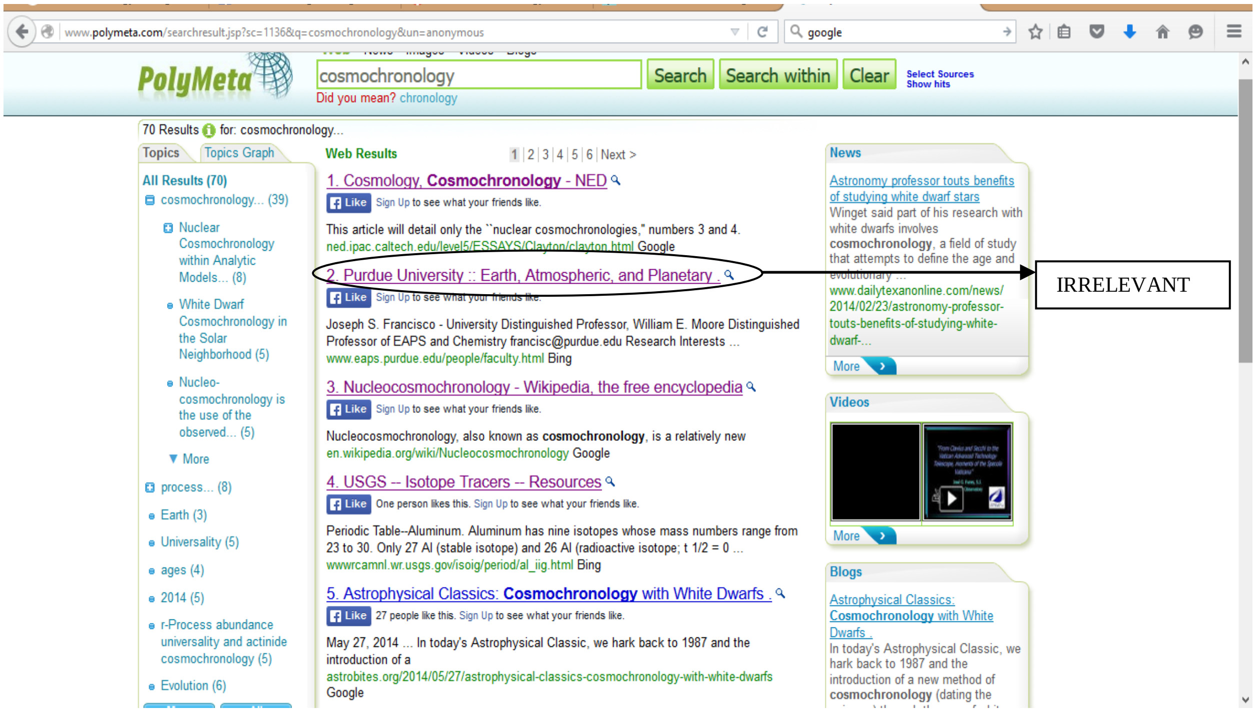
Cosmo-Chronology - Calendar Wizard
www.calendarwizard.org/cosmo-chronology.html
Quranic Hadheeth Ayats and Cosmo-Chronology. This brings into view our reckoning of time as related to our lives "here" and in the "hereafter".

Cosmopolitan - Facebook
www.facebook.com/Cosmopolitan/timeline
See more of Cosmopolitan by logging into Facebook. Message this Page, learn about upcoming events and more. If you don't have a Facebook account, ...

Neutron reactions and nuclear cosmo-chronology | Marcus ...
www.academia.edu/.../Neutron_reactions_and_nuclear...
Neutron reactions and nuclear cosmo-chronology. Authors. Marcus Dahlfors + 19. Marcus Dahlfors. Adonai Herrera-martinez. G. Vannini. M. Heil. C. Pretel. C. Guerrero ...

Nucleo-cosmochronology (Journal Article) | SciTech Connect
www.osti.gov/scitech/biblio/4409415
A review is made of the basis of nucleo-cosmochronologies, emphasizing the model independent approach. The exponential model as well as recent galactic evolution ...

Figure 21: InfoSpace Results



Did you mean? chronology

70 Results for: cosmochronology...

All Results (70)
- cosmochronology... (39)
  - Nuclear Cosmochronology within Analytic Models... (8)
  - White Dwarf Cosmochronology in the Solar Neighborhood (5)
  - Nucleo-cosmochronology is the use of the observed... (5)
  - More
- process... (8)
- Earth (3)
- Universality (5)
- ages (4)
- 2014 (5)
- r-Process abundance universality and actinide cosmochronology (5)
- Evolution (6)

Web Results

1. Cosmology, **Cosmochronology** - NED
Sign Up to see what your friends like.
This article will detail only the ``nuclear cosmochronologies," numbers 3 and 4.
ned.ipac.caltech.edu/level5/ESSAYS/Clayton/clayton.html Google

2. Purdue University :: Earth, Atmospheric, and Planetary .
Sign Up to see what your friends like.
Joseph S. Francisco - University Distinguished Professor, William E. Moore Distinguished Professor of EAPS and Chemistry francisc@purdue.edu Research Interests ...
www.eaps.purdue.edu/people/faculty.html Bing

3. Nucleocosmochronology - Wikipedia, the free encyclopedia
Sign Up to see what your friends like.
Nucleocosmochronology, also known as **cosmochronology**, is a relatively new
en.wikipedia.org/wiki/Nucleocosmochronology Google

4. USGS -- Isotope Tracers -- Resources
One person likes this. Sign Up to see what your friends like.
Periodic Table--Aluminum. Aluminum has nine isotopes whose mass numbers range from 23 to 30. Only 27 Al (stable isotope) and 26 Al (radioactive isotope; t 1/2 = 0 ...
wwwrcamnl.wr.usgs.gov/isoig/period/al_iig.html Bing

5. Astrophysical Classics: **Cosmochronology** with White Dwarfs .
27 people like this. Sign Up to see what your friends like.
May 27, 2014 ... In today's Astrophysical Classic, we hark back to 1987 and the introduction of a
astrobites.org/2014/05/27/astrophysical-classics-cosmochronology-with-white-dwarfs Google

News

Astronomy professor touts benefits of studying white dwarf stars
Winget said part of his research with white dwarfs involves **cosmochronology**, a field of study that attempts to define the age and evolutionary ...
www.dailytexanonline.com/news/2014/02/23/astronomy-professor-touts-benefits-of-studying-white-dwarf-...

Videos

Blogs

Astrophysical Classics: **Cosmochronology** with White Dwarfs .
In today's Astrophysical Classic, we hark back to 1987 and the introduction of a new method of **cosmochronology** (dating the

Figure 22: PolyMeta Results

The performance of MetaFusion is compared with existing MSEs i.e. Dogpile, InfoSpace and PolyMeta in terms of precision. The precision value of our proposed MetaFusion model comes out to be 0.624 whereas the precision value of Dogpile is 0.588 , InfoSpace is 0.521 and PolyMeta is 0.500. Hence the precision of proposed model MetaFusion is greater than existing research MSEs which is shown in Figure 23.

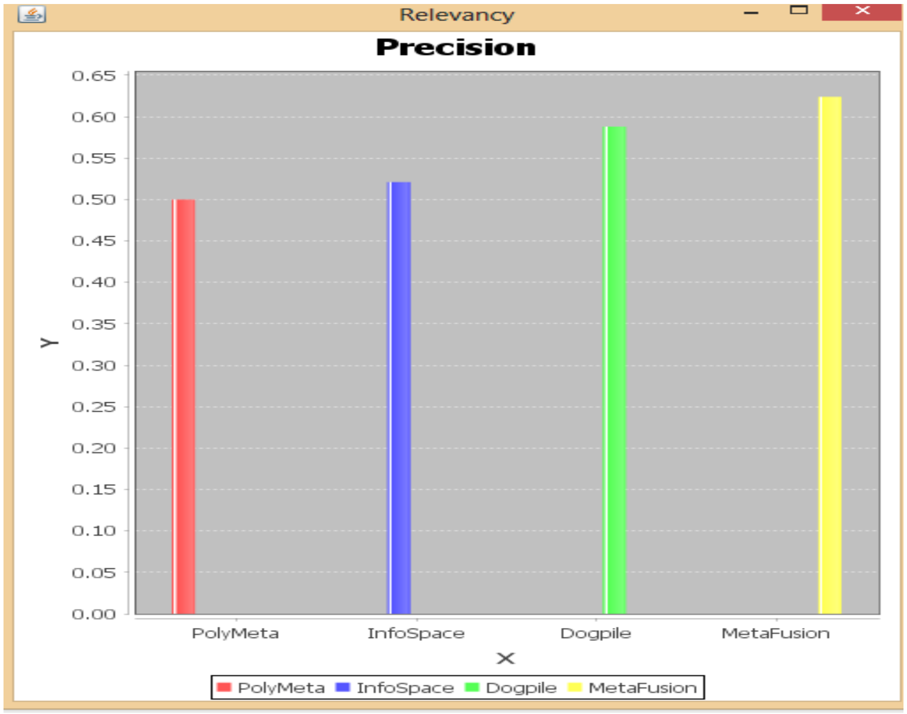


Figure 23: Comparison of Precision

5.3 Comparative Study

The comparison is performed between our proposed model MetaFusion and several existing models. Table 4 presents the comparison.

Table 4: Comparison of MetaFusion with other models

Evaluation criteria	**MetaFusion**	**MetaXplorer**	**MetaSurfer**	**OWA model**
Adapts with changing environment	Yes	Yes	No	No
Underlying techniques used	FAHP and Genetic	FAHP and In-OWA operator	FAHP and modified EOWA	OWA operator

	Algorithm		operator	
Expert free decision making process	Yes (search engines are assigned weights by applying learning algorithm)	Yes (importance degree are learned)	No (experts assigns weight to search engines)	No
Multi-optimization algorithm	Genetic algorithm performs optimization	No	No	No
Performance evaluation against research MSEs	Performs better than Dogpile, Infospace and Polymeta over set of 100 queries	Performs better than Excite and Webcrawler over set of 30 queries	Performs better than Mamma, Excite and Webcrawler over set of 14 queries	No

CHAPTER 6: CONCLUSION AND FUTURE WORK

This chapter discusses the conclusions drawn from our research work and also presents the future work which can be done in this research .

6.1 Conclusion

World Wide Web contains enormous number of documents and is major source of information dissemination. It is very challenging task to retrieve relevant set of documents from large database. Hence our proposed meta-search engine model ,MetaFusion , tries to retrieve the relevant information according to user query. The previous model of MSE, MetaXplorer uses FAHP for result aggregation whereas our proposed algorithm uses Fuzzy AHP along with Genetic algorithm to generate the aggregated rank list and arrange the documents in order of their decreasing fitness value i.e. “survival of fittest”. Hence the document at the top will be having higher fitness value. In our proposed work Genetic algorithm is applied for multi-optimization and thus it is more efficient. The performance of MetaFusion is compared with available research meta-search engines over set of 100 test queries which are taken evenly from different research domains. The results shows that MetaFusion has the highest precision of 0.624 when compared with available research MSEs Dogpile , Infospace and PolyMeta.

Hence ,the major advantages of the proposed model, MetaFusion can be summarized as:

i. MetaFusion is adaptable and dynamic in nature because training algorithm can be run periodically to reflect environmental changes.

ii. MetaFusion does not depend on biased opinions of decision maker in assigning importance degrees to search engine.

iii. In Earlier models , experts manually inspects search engine’s performance but in our proposed model, MetaFusion , importance degree of search engine’s are learned automatically in training phase.

iv. In Earlier models only pre-processing and post-processing of results happen , but in our proposed model internal processing i.e. URL Analysis also happens along with pre and post processing.

v. Proposed model, MetaFusion is adaptable to changing environment because if users are not satisfied with results ,they can provide their feedback. This in turn would cause training phase to be executed again , if a considerable number of user requests have been received.

vi. Proposed model uses Fuzzy AHP along with Genetic Algorithm to generate optimized results.

vii. Proposed model is flexible and robust because uses Genetic Algorithm in result merging.

6.2 Future work

In future , we can extend this research work by considering more number of search engines in result aggregation. When we add more number of search engines, number of retrieved documents increases. This will increase the count of retrieved relevant and irrelevant documents. Also number of duplicate documents will increase. Hence, we need a measure for duplicate removal and appropriate technique should be applied to limit the number of finally displayed documents. Also in future , we can apply any other multi-optimization algorithm for result merging.

CHAPTER 7: PUBLICATIONS FROM THE RESEARCH

This chapter briefly states the communicated research paper from this research work , along with details of the conference of publication.

1. Daya Gupta , Devika Singh, **“ *MetaFusion: An Efficient MetaSearch Engine using Genetic Algorithm*” ,** 9th International Conference on Contemporary Computing(IC3 2016), IEEE.

REFERENCES

[1] R.M. Losee, “When information retrieval measures agree about the relative quality of document rankings”, Journal of the American Society of Information Science, vol. 51, issue 9, pp. 834-840, 2000.

[2]Keyhanipour, A. H., Moshiri, B., Kazemian, M., Piroozmand, M., and Lucas, C., “ Aggregation of Web search engines based on users’ preferences in WebFusion”, Knowledge-Based Systems, vol. 20, issue 4, pp. 321–328, 2007.

[3] Bar-Ilan, J., Mat-Hassan, M., and Levene, M., “Methods for comparing rankings of search engine results”, Computer Networks, vol. 50, pp. 1448–1463, 2006.

[4] Spink, A. H., Jansen, B. J., Blakely, C., and Koshman, S., “A study of results overlap and uniqueness among major Web search engines”, Information Processing & Management, vol. 42, issue 5, pp. 1379–1391, 2006. 32

[5] Spoerri, A., “Examining the authority and ranking effects as the result list depth used in data fusion is varied”, Information Processing & Management, vol. 43, issue 4, pp. 1044–1058, 2007.

[6] Vaughan, L., “New measurements for search engine evaluation proposed and tested”, Information Processing & Management, vol. 40, issue 4, pp. 677–691, 2004.

[7] Available Online at: http://en.wikipedia.org/wiki/Metasearch_engine

[8] A. Gulli and A. Signorini, The Indexable Web is more than 11.5 billion pages. In Poster proceedings of the 14th international conference on World Wide Web, ACM Press, pages 902--903, Chiba,Japan, 2005.

[9] Manoj M., Elizabeth Jacob, “Information Retrieval on Internet using meta-search engines: A review”, Journal of Scientific and Industrial Research, vol. 67, pp. 739-746, 2008.

[10] Weiyi Meng, Clement Yu and King-lup Liu, “Building Efficient and Effective Metasearch Engines”, ACM Computing Surveys, vol. 34, issue no. 1, pp. 48-89, 2002.

[11] Erik Selberg, Oren Etzioni, “Multi-Service Search and Comparison Using the MetaCrawler”, Proceedings of the 4th International World Wide Web Conference, pp. 195-208, 1995.

[12] Erik Selberg, Oren Etzioni, “The MetaCrawler Architecture for Resource Aggregation on the Web”, IEEE Expert, 1997.

[13] J. Aslam and M. Montague, "Models for Metasearch", Proceedings of the 24th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, LA, USA, pp. 276-284, 2001.

[14] E. D. Diaz, A. De, V.V. Raghavan, “A Comprehensive OWA based Framework for Result Merging in Metasearch”, 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, Canada, Springer, pp. 193-201, 2005.

[15] Devendra Tayal, Amita Jain, Neha Dimri, Shuchi Gupta, “MetaSurfer: a new metasearch engine based on FAHP and modified EOWA operator”, International Journal of System Assurance Engineering and Management, Springer, 2014.

[16] Daya Gupta, Neha Dimri, “ MetaXplorer: An Intelligent and Adaptable Metasearch engine using a novel OWA operator” ,Mtech Thesis, DTU 2015.

[17] Holland J H, Adaptation in natural and artificial systems, MIT Press, 1975.

[18] Available Online at: http://www.alexa.com/

[19] C. C. Vogt., “How much more is better? Characterizing the effects of adding more IR systems to a combination”, Content-Based Multimedia Information Access (RIAO), Paris, France, pp. 457–475, 2000.

[20] Ronald R. Yager, “On Ordered Weighted Averaging Aggregation Operators in multicriteria Decisionmaking”, IEEE Transactions on Systems, Man, and Cybernetics, vol. 18, pp. 183-190, 1988.

[21] Christer Carlsson, Robert Fullér, and Szvetlana Fullér, “OWA Operators for doctoral student selection problem”, R.R.Yager and J.Kacprzyk eds., The ordered weighted averaging

operators: Theory, Methodology, and Applications, Kluwer Academic Publishers, Boston, pp. 167-178, 1997.

[22] Torra, V., “OWA operators in data modeling and re identification”, Fuzzy Systems, IEEE Transactions, vol. 12, Issue 5, pp. 652 – 660, 2004.

[23] José M. Merigó and Anna M. Gil-Lafuente, “Using the OWA Operator in the Minkowski Distance”, International Journal of Social and Human Sciences, vol. 2, 2008.

[24] José M. Merigó, Anna M. Gil-Lafuente, “Decision Making with the OWA operator in sport management”, Expert Systems with Applications, Volume 38, Issue 8, pp. 10408–10413, 2011.

[25] Arijit De, Elizabeth E. Diaz and Vijay V. Raghavan, “On Fuzzy Result Merging for Metasearch”,Fuzzy System Conference, IEEE, pp. 1-6, 2007.

[26] Francisco Chiclana, Francisco Herrera, and Enrique Herrera-Viedma, “The Ordered Weighted Geometric Operator: Properties and Application in MCDM problems”, in Proc. 8th Conference on Information Processing and Management of Uncertainty in Knowledge based Systems (IPMU), pp. 985-991, 2000.

[27] Azcel, J., and Alsina, C., “Procedures for synthesizing ratio judgments”, Journal of Mathematical Psychology, vol. 27, pp. 93-102, 1983.

[28] Azcel, J., and Alsina, C., “Synthesizing judgements: A functional equations approach”, Mathematical Modelling, vol. 9, pp. 311-320, 1987.

[29] F. Chiclana a, E. Herrera-Viedma b, F. Herrera b, and S. Alonso, “Some induced ordered weighted averaging operators and their use for solving group decision-making problems based on fuzzy preference relations”, European Journal of Operational Research, vol. 182, pp. 383–399, 2007.

[30] Zarghami, M., Ardakanian, R., Memariani, and A., Szidarovszky, F., “Extended OWA operator for group decision making on water resources projects”, Journal of Water Resources Planning and Management, vol. 134, issue 3, pp. 266–275, 2008.

[31] M.Q. Suo, Y.P.Li, G.H.Huang, “Multicriteria decision making under uncertainty: An advanced ordered weighted averaging operator for planning electric power systems”, Engineering Applications of Artificial Intelligence, vol. 25, issue 1, pp. 72-81, 2012.

[32] E. D. Diaz, “Selective Merging of Retrieval Results for Metasearch Environments”, Ph.D. Dissertation, University of Louisiana, Lafayette, LA, 2004.

[33] T. L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, New York, 1980.

[34] T.L. Saaty, "Relative Measurement and its Generalization in Decision Making: Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors - The Analytic Hierarchy/Network Process". Review of the Royal Spanish Academy of Sciences, Series A, Mathematics, vol. 102, 2, pp 251-318, December 2007.

[35] Dinesh. M.S, K.ChidanandaGowda and P.Nagabhushan, “Fuzzy Hierarchical Analysis for Remotely Sensed data”, Geoscience and Remote Sensing Symposium Proceedings, IEEE, vol. 2, pp. 782-784, 1998.

[36] M.H. Vahidnia, A. Alesheikh, A. Alimohammadi, and A. Bassiri, “Fuzzy Analytical Hierarchy Process In GIS Application”, The International Archives of the Photogrammetry, Remote Sensing and patial Information Sciences, Vol. 37, Part B2, Beijing, 2008.

[37] Gwo-hshiung Tzeng, Min-Jiu Hwang, Jia-Horng Shieh, and Hsin-Chi Wu, “Applying Fuzzy AHP and Nonadditive Fuzzy Integral Methods for Evaluation and Selection of Construction Project Contractor”, 6th ISAhP, 2001.

[38] Feng Kong, Hongyan Liu, “Applying Fuzzy Analytic Hierarchy process To Evaluate Success Factors of E-Commerce”, International Journal Of Information and System Sciences, vol. 1, pp. 406-412, 2005.

[39] Yu-Cheng Tang and Malcolm J. Beynon, “Application and Development of a Fuzzy Analytic Hierarchy Process within a Capital Investment Study”, Journal of Economics and Management, vol. 1, pp. 207-230, 2005.

[40] Yumei Chen, “Fuzzy AHP-based Method for Project Risk Assessment”, Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010.

[41] Hing Kai Chan, “An Extended Fuzzy-AHP Approach for the Evaluation of Green Product Designs”, IEEE Transactions On Engineering Management, Issue 99, pp. 1-13, 2012.

[42] Arijit De, Elizabeth Diaz, “Hybrid Fuzzy Result Merging for Metasearch Using Analytical Hierarchy Process”, 28th North American Fuzzy Information Processing Society Annual Conference (NAFIPS), USA, IEEE, 2009.

[43] SUN Ying-cheng ,LI Qing-shan, “The Research Situation and Prospect Analysis of Meta-search Engines 2012”,International Conference on Uncertainty Reasoning and Knowledge Engineering,IEEE 2012.

[44] Ming Zhou, Shudong Sun, Basic principle and application of genetic algorithm. National Defense Industry Press, Beijing, 1999.

[45] Srinivas, N., Deb, K., Multiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary Computation, 2(3), 221-248.

[46] Waqas, K.; Baig, R.; Ali, S., "Feature subset selection using multiobjective genetic algorithms," *Multitopic Conference, 2009. INMIC 2009. IEEE 13th International* , vol., no., pp.1,6, 14-15 Dec. 2009.

[47] Jinho Kim; Zong Woo Geem, "Optimal scheduling for maintenance period of generating units using a hybrid scatter-genetic algorithm," *Generation, Transmission & Distribution, IET* , vol.9, no.1, pp.22,30, 1 8 2015.

[48] Biswas, K.; Muthukkumarasamy, V.; Singh, K., "An Encryption Scheme Using Chaotic Map and Genetic Operations for Wireless Sensor Networks," *Sensors Journal, IEEE* , vol.15, no.5, pp.2801,2809, May 2015.

[49] Shih-Chia Huang; Ming-Kai Jiau; Chih-Hsiang Lin, "A Genetic-Algorithm-Based Approach to Solve Carpool Service Problems in Cloud Computing," *Intelligent Transportation Systems, IEEE Transactions on*, vol.16, no.1, pp.352,364, Feb. 2015.

[50] M. Gordon. “Probabilistic and Genetic Algorithms in Document Retrieval” Commun. ACM, 31(10):1208–1218, 1988.

[51] K. Deb and S. Agrawal. A niched-penalty approach for constraint handling in genetic algorithms. In Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA-99), pages 235–243. Springer-Verlag, 1999.

[52] Available Online at: https://developers.google.com/custom-search/

[53] Available Online at: https://code.google.com/p/matlabcontrol/

[54] Available Online at: http://www.java2s.com/Code/Jar/h/Downloadhttpclient41jar.htm

[55] Available Online at: http://www.java2s.com/Code/Jar/h/Downloadhttpcore41jar.htm

[56] Available Online at: http://www.java2s.com/Code/Jar/h/Downloadhttpmime41jar.htm

[57] Available Online at: https://datamarket.azure.com/dataset/bing/search

[58] Available Online at: http://commons.apache.org/proper/commons-logging/

[59] Available Online at: https://commons.apache.org/proper/commons-codec

[60] Available Online at: https://commons.apache.org/proper/commons-net/download_net.cgi