Research Journal of Applied Sciences, Engineering and Technology 13(6): 432-441, 2016
DOI:10.19026/rjaset.13.3003
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: May 25, 2015 Accepted: July 26, 2015 Published: September 15, 2016

Research Article

Study on Search Engine Optimization Technique in Health Care System using Cloud Data

[1, 2]R. Kavitha and [3]R. Nedunchelian
[1]Saveetha School of Engineering, Saveetha University,
[2]Department of Computer Science and Engineering, Bharath University,
[3]Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Chennai, Tamil Nadu, India

Abstract: Normally Search Engine System will work, based on web crawling, text classification and information retrieval. With the rapid growth of network information resources, more and more people are concerned about how quickly and efficiently the communication from the mass of network information extraction required. The Search Engine Technology is one of the main means of achieving network information mining. This study analyzes basic search engine of these defects on the combination of the structural characteristics of search engine optimization gives a specific method, the optimization method can overcome the existing search. Corresponding engine inherent defects, improve the effectiveness of search engines at the same time, to further promote the intelligent search engine. The entire search engines are focused Keywords, Reserved words, Word frequency, Weightage. This study focuses health care search Engine system efficiency; Using C-means with Neuro-Fuzzy System, the search engine based on intelligence provides improved services in the health care when compared with other search engines.

Keywords: C-means, neuro fuzzy system, search engine

INTRODUCTION

Even though search engines have good performance, there are some flaws like slow speed, dead links and duplicate or irrelevant information. The information that they throw need not necessarily be the user's desired content and also they may be too many results to check. The sites can also be irrelevant. This poses a great challenge to the validity and reliability of the search engines. Existing search engines do not meet the user's expectations in terms of quality and meeting higher volume levels with faster access for a variety of information required.

Hence, it is paramount to address this issue as people need relevant information and services without wasting much of their precious time. Effective optimization has become an important part of search engine research. The overall goal is to make search engines more intelligent and precise. SEO increases the search rank of the websites. This is possible by adding more contents that are relevant to the keywords searched by the users. User's mindset, habits and the way that they search for information is also important.

Specific methods to overcome this defect in accordance with the following steps:

First, when the user searches for information, the words spread on the internet in order to find and search for the information which could be the discovery and introduction of a collection of letters. There is a control parameter for the amount of information retrieved which is known as the threshold value. When the information collected reaches this value, the user cannot continue to operate. So, it is not necessary that to obtain higher call rate, the endless search should continue. Therefore, the number of queries from the source for information can be controlled effectively.

Secondly, Indexing paves path directly to the retrieval of information effectively. Speed and performance can be optimized with the help of indexing the relevant document; else the search engine could search every possible document. Indexing has two parameters, Exhaustive and non-exhaustive which controls the effectiveness of search engines. In Exhaustive indexing, terms are generated in large numbers pertaining to the matter present in the document. In Non-Exhaustive indexing, fewer terms are generated that are relevant to the documents. The Chinese index is based on the Chinese language. Based on the natural language, the technical contents and picture contents are understood. In particular, automatic classification like energy technology, clustering technology, automatic summarization technology, machine learning technologies can be applied to determine the search of information. Samples can be

Corresponding Author: R. Kavitha, Saveetha School of Engineering, Saveetha University, Chennai, Tamil Nadu, India

implemented in the index section. This effectively reduces the duplication of redundant information.

Third, In accordance with the crawler, the function is to retrieve the time. The query household's the rapid detection information document library in the index. The documentation of related query evaluation can be given a quantitative correlation meter. The Operator program (which has a lot of achievements in technology, such as text classification (Bhaskaran *et al*., 2013; Arzanian *et al*., 2010)), is used to obtain the correlation coefficient. Depending on the size of the correlation coefficient, the index documents in the library can be accorded the correlation coefficient based on the descending sort program. The limit is set to the length of the parameter. To achieve this, the parameter value will no longer continue to sort the information and discard the text.

LITERATURE REVIEW

There exists a need for Multimedia information systems and Web-based applications like search engines to extract and classify data automatically. A framework to formulate the similarity-based queries is proposed. Here, we define a fuzzy relational algebra which permits in formulating queries related to the similarity of the objects. Equivalence and containment rules appropriate for optimizing queries are also studied. The focus will be on developing more complex data models in the near future (Montesi and Trombetta, 1999).

An annealing algorithm based on Internet agent simulated in hybrid is used for searching home pages that are relevant on WWW. The agent searches for more relevant home pages after obtaining a set of home pages and can operate autonomously without human intervention. By comparing the finest-first search with the annealing algorithm which is hybrid simulated, a higher fitness score is achieved. A higher precision and recall can also be obtained (Yang *et al*., 2000).

A new architecture and algorithm for Directory services is considered to optimize the searching performance. For this to succeed, signature and n-gram method are combined to improve the filtering accuracy and performance. Since the stored data are more or less short strings in Directory servers, the properties of short strings should be considered by the estimated search engines for service in the Directory. As a result, the signature algorithm was found to be better than the original algorithm (Yang *et al*., 2003).

In recent years, search engine marketing industries are growing at a rapid pace. At present, it is categorized as paid and non-paid. Non-paid marketing is relatively new to the market and many companies do not have the awareness or means to handle their promotion tactics and make the most of their investments. The effect of this key dimension, reach and variety and to devise a Search Engine Marketing Management System (SEMMS) based on web are investigated. Using SEMMS, largely used keywords to be added can be identified in contents of the website. Development of SEMMS helps in identifying appropriate search engine tools and provides essential information to take better marketing decisions (Yu and Lin, 2007).

Text mining is relatively a new field which flanks on mining of data, machine learning, information retrieval, statistics and computational linguistics. Some of the Test mining operations are search and retrieval, feature extraction, text-based navigation, categorization and clustering. A review on Text Mining methods and applications are conducted. Since most of the information is stored in the text form, mining of text is supposed to have high commercial value (Gupta and Lehal, 2009).

Generally, search engine rankings are shaped by search engine users, companies and developers, optimization practitioners. Aspects like audiences of web content and competitors of website, keywords that appear on search engine result pages and involving web contents with other websites and web creators should be considered always. Professional communicators should apply these features and be updated with the evolving search algorithms, SEO practices, their website's traffic and their competition (Killoran, 2013).

Online discovery learning is well supported by digital repository on the web. Recently, the flexibility of World Wide Web has increased to a great extent due to improved infrastructure, resource learning and their ways of sharing. But e-learners find it very hard to retrieve information that relevant and specific to location. The amount of time spent on searching is more than that of learning. Multiple agents delivering personalized SERP is used and is more suitable for personalizing web pages. This also uses mining techniques with web usage to establish the knowledge of the learner (Shah and Jain, 2011).

Search Engine Optimization (SEO) techniques are applied in the tourism industry. To suggest the specific process of SEO, a motel website is taken for example. The results indicate that after applying the process of CEO, the ranking and bandwidth have increased notably in the motel website. The telephone enquiries and accommodation amount shows better marketing effect when compared without SEO process and also shows the seasonal preferences of the customer. Therefore, it is obvious that we can apply SEO internet marketing strategy to the tourism industry (Ho *et al*., 2010).

A broad search volume based keywords evaluation method is suggested. Based on the search volume, a more simple and systematic way of selecting keywords can be identified by deciding the proper keywords as the target of optimization using the complex index evaluation method. The method is found to be practical and efficient, though considering only the part selection of the suitable keyword search volume (Gao *et al*., 2010).

Search Engine Optimization (SEO) techniques are normally monitored from four features, namely structure, keywords, content and link optimization. The

focus is on how the optimization techniques can have impact on the efficiency of SEO and Page interest. With practical experience and literary review, the model is constructed and hypothesized. To assess the SEO techniques, intermediary measuring tool is used on the data collected from various websites. The results indicate that the Page Size has a notable positive effect on both efficiency of SEO techniques and Page Interests (Wang, *et al*., 2011).

The contest between the Search Engine Provider's (SEPs) Paid Search Marketing (PSM) and Search Engine Optimization firm's Search Engine Optimization (SEO) is a fiercely fought one and is growing at a rapid pace. At present, SEP leads the search engine advertising market. With better algorithm efficiency, search engines can boost the confidence of both PSM and SEO (Li *et al*., 2014).

The index freshness and coverage, the problem of capacity and buffer size determination of ER is scrutinized to optimize two objectives of conflict together. To solve this problem, an approach based optimization model is considered to determine an optimal policy in the theory of inventory control. The experiments confirm that this model attains low average cost than other models. After employing this model as its ER policy, it was observed that Feedmil.com was able to obtain fresh search results (Shin *et al*., 2012).

Management can use Business Intelligence (BI) or business analytics tools to obtain exact and timely information to lead a healthcare organization. But, huge capital investments are required for implementation. Hosted in the cloud computing environment, the numerous shades of healthcare business analytics are evaluated. To enhance the functions in a healthcare enterprise, BI is presented as Software as a Service (SaaS) solution. This solution also helps in identifying the problems faced by the healthcare enterprises (Bhattacharya *et al*., 2012).

Healthcare practitioners don't have complete access to the medical information about a patient, which is essential for them to make timely clinical decisions. A Cloud-based solution that offers a consolidated patient-relevant data to health care providers in the Indian context is looked upon. This solution has been deployed at a multi-site healthcare facility. Based on the solution it is believed that the eccentric cloud-based architectures promise to address data integration problems in the context of developing countries (Bhaskaran *et al*., 2013).

In emergency medical treatment, Emergency Medical System (EMS) is a radical approach in some medical emergency. Using cloud computing, it explains a mobile system that allows electronic healthcare data storage, update and retrieve the data. An Android Based Tracking on cloud for Emergency Medical System (EMS) can locate the nearest hospital, ambulance emergency contact system, accessing Electronic Health Record of emergency patient and can aid in pre-hospital treatments (Nimbalkar and Fadnavis, 2014).

Search engine optimization can be benefitted, if the keywords combinations are identified by utilizing TF-IDF technologies, K-means clustering and indexing quality examination. The combination of keywords can then be retrieved effectively for benefitting the website. Web sites of small and medium scale enterprises can increase the website's exposure level for successful operation of website through search engine optimization (Lin and Chi, 2014).

Based on Heuristic Semantic Walk (HSW), a randomized approach to search a collaborative network in order to extract meaningful semantic chains between concepts is undertaken. Major problems posed by collaborative networks like large dimensions, high connectivity degree and dynamic evolution of online networks which makes traditional search methods ineffective and not viable is solved by this method. The additional potential applications range also includes user navigation simulation, argumentation mining and query expansion (Franzoni *et al*., 2014).

A design for integrating Information and Informatics framework that permits to store, integrate and analyze healthcare data in a cloud-based system is evaluated. This framework will allow the development of latest healthcare application with data integrated in a different database. So, developers can quickly develop healthcare applications without thinking about data management in the cloud. These advanced healthcare applications will enhance the required data integration and facilitates faster interaction between service providers and the patients. Cloud platform Aneka is used for development (Sultana *et al*., 2014).

Cloud computing infrastructure is implemented to access various health care services over the internet in the healthcare industry. The key advantage of this computing is that it can act as catalyst for various entities in the healthcare field. By using a query-based optimal medication framework for better health care services with reasonable response time, the queries in natural language through a user interface and through first-order logic based evaluation engine it processes the input is accepted. A zChaff SAT solver is used to build the query evaluation engine (Majhi and Bera, 2014).

A new Fuzzy System Modelling with IFF approach for modelling systems with continuous output variable and uses a new Improved Fuzzy Clustering (IFC) algorithm is aimed at addressing some of the challenges like complex computation, subjectivity, optimization issues, etc., faced by traditional fuzzy models. IFC is used to fuzzy partition the data and uses enhanced membership values to establish fuzzy functions for every cluster. The system behavior was found to be more accurate when compared to other FSM models (Celikyilmaz and Burhan Turksen, 2008).

To implement an Ontology-based Web Content Management System (OWCMS), a Collaborative Ontology Learning Approach is applied. Content-based and User-based Learning are the two supervised learning approaches integrated. To extract ontology

concepts, text mining methods are applied in the Content-based Learning Approach and an Ontology Graph (OG). The feature analysis method is applied to User-based Learning Approach to pull out the rift of the Ontology Graph. The two methods are combined by the system to formulate mutual ontology learning via an ontology matching and refinement process (Lim *et al*., 2009).

To gather the results of numerous Chinese search engines, attribute properties of the web page and behaviour properties in a common search engine are the two main aspects considered. Data collected from the search engines are ranked again through comparing the scores in weight scheme. The transfer of returned data takes place and is stored in the virtual XML document. The research results show that this scheme is highly essential and also quite reasonable to rank the relative web page (Zhang, 2009).

MATERIALS AND METHODS

Comparison of key roles of fuzzy c-means algorithm and page ranking algorithm in search engine optimization: Some of the keywords are extracted from the browsing history of the user using FCM cluster. According to that, it will find the user's most searching keywords. Clustering the user into some clusters in which groups are formed by using the browsing activity of the user, using neighbours based on the collaborative filtering mechanism. Here, weight-age is given according to the individual user. The weight-age cluster of same group people is identified as per his/her action during the browsing.

The main aim is to reduce the searching time of the user via Search Engine Optimization. Because, most of the researcher spends his time searching for some concepts, ideas, etc. Fuzzy C- Means Clustering Algorithm and Page Rank algorithm play an important role in this assignment. Fuzzy C-Means clustering algorithm is used to retrieve the content regarding the keywords and Page ranking is used to rank the page using term weight-age computing ranking positions and the slow moving frequency of websites.

Page Rank: Page Rank is the nothing but the number value assigned by the search engine. Various search engines follow different calculation. Popular sites including Google, use this algorithm. But during peak hours it can't work. Originally most of the browsers can't support.

Damping factor: In accordance with the Page Rank Theory, a fantasy surfer may stop clicking on links finally even though he is causally clicking. The likelihood is that the individual will carry on is a damping factor d. It is presumed that the damping factor would be set approximately at 0.85. The damping factor is deducted from 1 and to the damping factor's product, this is added. It means the PageRank of pages is a derivative of the PageRanks from other pages. The damping factor slides the derived value. The next formula maintains the original report in Page and Brin's paper which proclaims "the sum of all PageRanks is one". According to Page and Brin's formula, a surfer is not interested after numerous clicks and moves to a random page. The PageRank value of a page replicates the chance that by chance the surfer will visit that page by clicking on the link. If there is no link between the pages, the random surfing process will get terminated. Nevertheless, there is a simple solution. If the random surfer visits a page which is in sink, it selects a different URL randomly and will carry on surfing. In PageRank calculation, pages without external links are believed to link all the pages in the collection. The PageRank scores are segregated among all 2010 pages evenly. These transitions are included in the Web to all nodes, with probability d = 0.85, which is a ballpark figure that a surfer uses the bookmark feature of the browser. The equation is : wherein p1, p2, ..., pN are the pages under consideration, M(pi) is the compilation of pages that relate to pi, L(pj) is the number of links that are outbound on page pj and N is the number of total pages. The entries of the eigenvector are the values of the PageRank of the adjacent matrix. This makes PageRank specifically elegant metric, where R is the solution of the equation and the adjacency function is 0 if page pj does not link to pi and normalized such that, for each j.

Factors considered in both papers:

Age/history of the domain: Domain's age is a vital factor which can increase the Page Rank. There are numerous domains which are many years old and have higher rankings when compared with other website in similar category. Search Engines also depend on the history of the domain to identify any SPAM activity. Everything can be changed in SEO, but not the Age/History of any domain. As a result, search engines could depend on this factor to award Trust status to particular websites.

Back links quality: The quality of the back links doesn't really mean higher Page Rank links. Usually, a website with Page Rank 2 is more valuable than buying a back link from a Page Rank 4 website akin. Search engines can identify any unusual pattern in the back links of the website. If 60 back links are purchased with PR 3-4, then Google can find it with no difficulty. But these actions will not only lower the Page Rank but also the Trust Rank. This kind of action can increase the Page Rank, but Truat Rank will be lowered which finally decides the website ranking.

Content on the website: Numerous article websites have faced a sharp decline in their traffic on both Yahoo and Google. The reason is the duplicate articles

submitted to various websites by different webmasters and SEO. Many of these websites have seen duplicate penalty. Some say that duplicate penalty does exist, but in the form of having the duplicate content in the additional index. But, the reality is that if the substance goes into supplement index, then the Trust Rank will be lowered. The focus should be on producing good-quality content. If the content is good and informative, other webmasters will start linking to the website automatically.

Optimization history of the domain: It is likely that search engines store the optimization history of any domain to establish their Trust Rank. If the website purchased 60,000 back links in 2005, then Google will keep a report of that activity to find out the Trust Rank.

COMPARATIVE STUDY

To search for healthcare content, the artificial neural network approach *of a* Kohonen self-organizing assists in attaining limited results. In order to store files from weblogs or websites locally, the widespread method called Web crawling is used. For Pattern (Rule) based text Classifiers, parsing and stemming to each text document are applied and converted into one text arrays. Data mining techniques based on frequency and redundancy are used to extract words. Documents are grouped into related tiny documents by using a C-Means classification algorithm.

For each group, labeling of documents is given on the basis of similarity features. All the details can be retrieved from the internet by the system once the user enters the keyword. For searching keyword, Neuro-fuzzy system's algorithm is used. The Neuro-fuzzy system mines the keywords history process according to the user search data. From the history of prior process, the system recognizes the rational consideration automatically when the user logins (Kavitha and Nedunchelian, 2014).

The research computes the level of the user participation from various search situations. The clauses involve both search field and value. On the basis of patent's features and search approach, numerous variables and weights like Patent Number (PN), International Patent Classification (IPC), Inventors (IN), Assignees (AN), Key Phrase (KP), Industry (INDT) and Technology Type (TECH) (Trappey *et al*., 2012) are defined by the research. The definition of the variable and equivalent weights is explained below:

Operation complexity
= Search actions score with weight+view action score with weight
+Analyze actions score with weight + bookmark action score with weight
+Expert actions score with weight
Explore:
The Operation Complexity:

$$OC = SAS + VAS + AAS + BAS + EAS \text{ with Weight}$$

The C-means with Neuro Fuzzy System methodology search is on the basis of the investigation of users' behavior patterns based on the user earlier search platform. The study computes each user's manners according to definite action types and investigates the user grouping result. Also, the suggestion method removes the appropriate search based on the shared filtering mechanism. The suggested method aids users get meaningful search even as cost and time can be saved.

Assume UT0001 as heart disease, UT0002 as lung disease, UT0003 as brain cancer, UT0004 as oral cancer, UT0005 as HIV disease etc.

The Operation Complexity is very less in paper C-means with Neuro-Fuzzy System compares with Intelligent Recommendation Methodology. Normally, the search engines depend on Keywords, Reserved words, Word frequency, Weightage of the word. But the C-means with Neuro-Fuzzy System search methods provide user type will provide service. If the user log in based on the previous search result the systems will provide service to the user.

An engine controller with noisy fitness functions is optimized online by GAs. For online optimization of vehicle engine controller, a GA that has Memory-based Fitness Evaluation (MFEGA) is applied. The conventional method was outperformed not only in convergence speed, but also in solution precision when numerical experiments were conducted in an engine simulator. The performance can be improved by introducing Bayesian technique (Sano *et al*., 2000).

A FOGA framework for generating fuzzy ontology on uncertain information is developed to allow generating fuzzy ontology which is furnished incrementally with new instances. The suggested framework will be useful in constructing ontology from uncertain data as it can signify uncertain information and build a concept hierarchy of uncertain information automatically (Tho *et al*., 2006).

SOM algorithm is used for retrieving documents along with a training algorithm for organizing documents and as well as retrieving documents. A software module is built to conduct an efficient search on documents related to education. The time performance of learning algorithms resulted in efficient retrieval and searching of the documents (Drigas and Vrettaros, 2006).

Analyzing the present news search engines prove that the provisions are inadequate and results organization is not sensible. So, FCA technique is used in optimizing the personal news search engine. In news search service, the formal concept analysis theory is used and the users' background is built. This results in

optimizing the query keywords and image of search results and also enhances the value of current news search services. To decrease the calculation costs like restricting the number of returning pages and disused data filtration from the pages of the web, a few strategies were implemented (Liu *et al*., 2007).

Traditional search engines have shortcomings in retrieving the potentially relevant documents. So in order to overcome this, an intelligent search engine based on ontology, known as CRAB, plans to set up natural language tools to mine information from web documents automatically and handle RDF triples effectively. New instances are identified by searching context-profiles of samples as well as testing examples in CRAB ontology. The CRAB ontology is then populated with these instances and placed them appropriately in the concept hierarchy (Shi and Sun, 2008).

Retrieving quick and precision information from the internet is one of the biggest challenges. Kernel-based Fuzzy C-Means is used for optimizing search. The results returned by the search engine are applied to the clustering algorithm and the relevant clustering results along with the query terms are then modified according to click through data of users' which in turn optimizes the query results (Lin and Zhang, 2009).

Meta-search engines may be capable of increasing the coverage scope of the web, but also returns a large amount of irrelevant data. The problem can be addressed by personalizing meta-search engines by filtering the results as per individual users' interest. A multi-agent architecture is used for modifying meta-search engine. The retrieved meta-search engine results are reordered by the system according to the users' interest. The system's personalized meta-search results are more pertinent when compared to the collective results of the search engines (Arzanian *et al*., 2010).

In recent times, search engine optimization technologies have gained a lot of attention for publishing business information quickly in order to maintain higher rankings. New strategies are introduced for the name space of domain, website content, links and keywords, which makes search engine optimization more effective. Along with search engine technology development, new strategies for search engine optimization are going to be a continuous exploration (Yunfeng, 2010).

Search Engine Optimization (SEO) is a method to enhance the importance of websites. By using reverse engineering, a system is built to crawl all factors of thousand of web pages by itself. An analysis of the content is made on few top positions on search result pages from Google. And finally top five factors are derived for SEO. Even for various websites, the main factors will always be the same (Zhu and Wu, 2011).

The act of increasing the page rankings of undeserving websites and thereby misleading the search engine is search engine spamming. Some commonly used black hat techniques are characterized and also a new method is discussed to oppose those techniques by combining page rank algorithm with link-based spam detection. This approach helps in discovering the target page and traces down the whole graph that is accountable for spreading spam (Somani and Suman, 2011).

To the original PageRank sorting algorithm, a new parameter factor is added and is known as grey factor. This grey factor is added to the PageRank algorithm in order to optimize it. After the PageRank algorithm is optimized, the result is expected to be more reasonable and will make up for the deficiencies of the just considered static factors (Wang *et al*., 2013).

BioNLP (Biomedical natural language processing) is a helpful method for unlocking priceless information that is stored in data of text format for research or practice. Besides dealing with large amount of terms which are specific to a domain, a sturdy biomedical parser is required to replica the related grammar. A system to get a grammar of probability based on a training body which consists of semantic classes and strings of concept from the UMLS is evaluated. On evaluation, the grammar derived achieved a recall of 0.644, average cross-bracketing of 0.61 and a precision of 0.737, which exhibited improved performance than a semantic-null control grammar (Fan and Friedman, 2011).

Clients generally contract out their data to the cloud storage servers in cloud computing, in order to cut management costs. The data is encrypted to shield the privacy of the data that is outsourced. The issue of Authorized Private Keyword Searches (APKS) over encrypted data is addressed in cloud computing. A scalable, fine-grained authorization framework is proposed to restrict the disclosure of sensitive information. On HPE over encrypted data, two ideal solutions for APKS are proposed. The first solution will improve the search efficiency and the second one will improve the query privacy (Li *et al*., 2011).

WWW encompass a massive number of websites globally. Since new information is added every day, researchers find it very difficult to retrieve relevant information from the huge volume of data. So it is imperative to design a search engine which saves precious time of the researchers and scientists. A framework to design a multi-agent based search engine which can retrieve desired information in an efficient manner and also with least pointer clicks is designed. Though primarily designed for researchers and scientists, this engine could also benefit ordinary users. They can also act as a platform for web-developers and vendors since they are part of e-marketing (Chheena and Ahmed Khan, 2012).

In order to identify the best set of keywords from the search results of the Top-k Keyword Proximity

Table 1: Comparing C-means with neuro fuzzy system vs intelligent recommendation methodology

	Techniques	Methods/Algorithm	Future	Reason to decrease the result
Paper 1	An effective system for automatic intelligent based Search Engine in health care cloud data using C-means with Neuro Fuzzy System.	Crawling pattern based Stemming-porter Classification-fuzzy c-means Page display-page ranking algorithm	Also consider time taken in each page. Scrolling speed in that page At a time how many pages opened	Reserved words will very low
Paper 2	Application of Webpage Optimization for Clustering System on Search Engine-Google Study	Benefit the Search Engine Optimization by using TF-IDF technologies, K-means clustering and indexing quality examination to identify the keywords combination.	Page display-page ranking algorithm	Websites of small and medium scale enterprises can increase the website's exposure level for successful operation of website through search engine optimization

Table 2: OC-score (C-means with neuro fuzzy system) vs OC-score (intelligent recommendation methodology)

User type	OC-score (C-Means with neuro fuzzy system)	OC-score (Intelligent recommendation methodology)
UT0001	0.021	0.0240
UT0002	0.021	0.0280
UT0003	0.103	0.1230
UT0004	0.011	0.1020
UT0005	0.031	0.0940
UT0006	0.011	0.0211
UT0007	0.121	0.1310
UT0008	0.021	0.0240
UT0009	0.021	0.0280
UT0010	0.113	0.1230
UT0011	0.111	0.1320

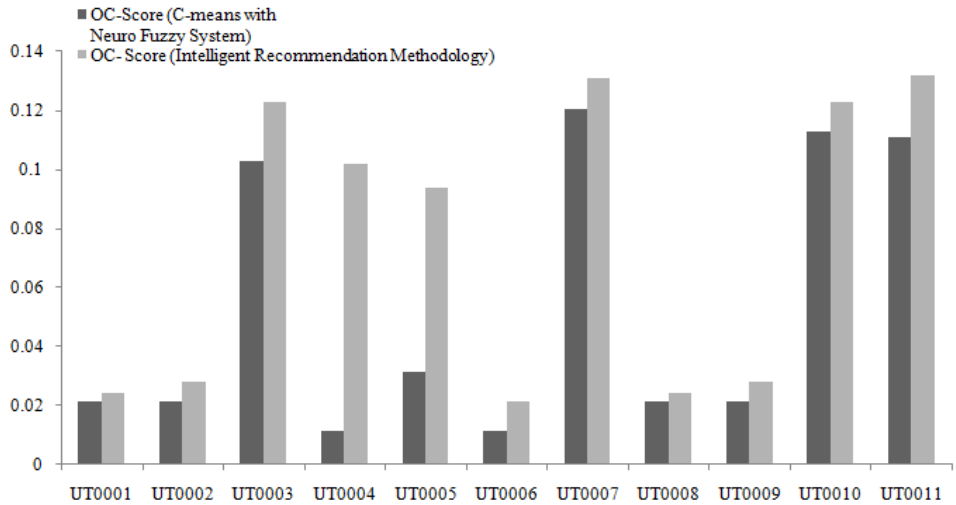

Fig. 1: OC-Score (C-means with neuro fuzzy system) Vs OC-score (Intelligent recommendation methodology)

Organic Searches, Similar Keyword Searches and Image Searches, a tactic is suggested. To find out the customers' likeability, demographics, demand patterns and trends, the database of the web services is analyzed. Internet of Things (IOT) has some differences with the active Internet for the people as the upcoming trends in the web server growth (Chung *et al*., 2012).

A semantic information retrieval system which is specific to domain is implemented based on ontology. This approach is unique as it uses the query concepts as well as its synonyms to carry out query expansion. Also, the new terms are only added if there is a similarity within a threshold and only the relevant documents will be ranked top. The application is more efficient after combining the technique of Automatic Query Expansion and semantic search (Chauhan *et al*., 2013).

A framework is used to develop a recommender system by combining an intelligent image retrieval and Information Retrieval (IIR) alongside user profile learning. In a university kind of setting, this framework was implemented to automate the physical search process of the answer booklets of examinations. The result showed that the recommender system was highly effective in retrieving information which is relevant to the topic or examination (Venkatraman and Kamatkar, 2013).

Feature Selection is an important means to improve classification. A set of classification algorithms is used to cover the latest learning schemes and evaluated using AUC, FPR and sensitivity. The results show that the AUC is the one of the efficient way for evaluating a classifier performance. The approach has enabled the usage of large datasets in known FR. The results are

best in AUC and FPR. It is evident that the correlated variables are reduced by FR (Santos *et al*., 2014).

Clustering is a process in which a set of objects is grouped and those groups have similar kind of objects. Lingo algorithm is used for this process. A cluster based searching which is introduced for XML keyword search, since XML can manage large amounts of data. This is very useful in retrieval of information application which also reduces the consuming time. This process involves designing of term-document matrix and extraction of frequently used phrases by using suffix arrays. The results achieved by Lingo are notably pragmatic (Ajitha and Gunasekaran, 2014).

For clustering web results, a new description-centric algorithm called IFCWR is introduced. Firstly, by using Forgy's tactic, IFCWR chooses a highest estimated number of clusters and combines cluster repeatedly until results are improved. To automatically evaluate a better solution, IFCWR uses the Bayesian Information Criterion. IFCWR was compared with other well-established web document clustering algorithms and the results showed a significant improvement in clustering performance and quality (Cobos *et al*., 2013).

For an efficient patent search, the method uses an automatic search engine. By clustering users' patent search pattern, the system gathers new patent proposal. Based on the collaborative filtering mechanism, the recommendation system extracts the suitable patents. This system is of great help to the users, as it not only obtains related patents but also helps in saving time and costs (Trappey *et al*., 2012).

Table 1 represents Comparing C-means with Neuro Fuzzy System Vs Intelligent Recommendation Methodology and Table 2 and Fig. 1 shows OC-Score (C-means with Neuro Fuzzy System) Vs OC- Score (Intelligent Recommendation Methodology).

CONCLUSION

This study discusses the techniques of the web content mining for retrieving the information effectively in the healthcare system. Normally, the World Wide Web (W3C) consortium is the dynamic environment that is the content of the data is changing day by day. Because of this problem, the user should not get the desired information effectively. The automatic intelligent based Search Engine in health care cloud data using C-means with Neuro-Fuzzy System solves this problem and helps the users to fulfill their needs. The Search Engine Optimization by using TF-IDF technologies, K-means clustering and indexing is used to retrieve the information effectively from the web and also it is used to improve the scalability. Therefore, the user can get the desired information efficiently.

REFERENCES

Ajitha, P. and G. Gunasekaran, 2014. Effective feature extraction for document clustering to enhance search engine using XML. J. Theor. Appl. Inform. Technol., 68(1).

Arzanian, B., F. Akhlaghian and P. Moradi, 2010. A multi-agent based personalized meta-search engine using automatic fuzzy concept networks. Proceeding of the IEEE 3rd International Conference on Knowledge Discovery and Data Mining (WKDD'10), pp: 208-211.

Bhaskaran, S., G. Suryanarayana, A. Basu and R. Joseph, 2013. Cloud-enabled search for disparate healthcare data: A case study. Proceeding of the IEEE International Conference on Cloud Computing in Emerging Markets (CCEM, 2013), pp: 1-8.

Bhattacharya, I., A. Ramachandran and B.K. Jha, 2012. Healthcare data analytics on the cloud. Online J. Health Allied Sci., 11(1).

Celikyilmaz, A. and I. Burhan Turksen, 2008. Enhanced fuzzy system models with improved fuzzy clustering algorithm. IEEE T. Fuzzy Syst., 16(3): 779-794.

Chauhan, R., R. Goudar, R. Sharma and A. Chauhan, 2013. Domain ontology based semantic search for efficient information retrieval through automatic query expansion. Proceeding of the International Conference on Intelligent Systems and Signal Processing (ISSP, 2013), pp: 397-402.

Chheena, N.D. and M.N. Ahmed Khan, 2012. Multi-agent based search engine for researchers and scientists. Int. J. Adv. Sci. Technol., 49(2012): 83-94.

Chung, P.T., S.H. Chung and C.K. Hui, 2012. A web server design using search engine optimization techniques for web intelligence for small organizations. Proceeding of the IEEE Long Island Systems, Applications and Technology Conference (LISAT, 2012), pp: 1-6.

Cobos, C., M. Mendoza, M. Manic, E. León and E. Herrera-Viedma, 2013. Clustering of web search results based on an iterative fuzzy c-means algorithm and bayesian information criterion. Proceeding of the Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), pp: 507-512.

Drigas, A.S. and J. Vrettaros, 2006. An intelligent search engine assessing learning material to improve learning procedures. Proceeding of the 7th International Conference on Information Technology Based Higher Education and Training (ITHET'06), pp: 875-883.

Fan, J.W. and C. Friedman, 2011. Deriving a probabilistic syntacto-semantic grammar for biomedicine based on domain-specific terminologies. J. Biomed. Inform., 44(5): 805-814.

Franzoni, V., M. Mencacci, P. Mengoni and A. Milani, 2014. Semantic heuristic search in collaborative networks: Measures and contexts. Proceeding of the IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 1: 141-148.

Gao, J.G., G. Yu and L. Tao, 2010. A method of keywords selection evaluation based on search volume. Proceeding of the International Conference on Educational and Network Technology, pp: 25-28.

Gupta, V. and G.S. Lehal, 2009. A survey of text mining techniques and applications. J. Emerg. Technol. Web Intell., 1(1): 60-76.

Ho, L.H., M.H. Lu, J.C. Huang and H.Y. Ho, 2010. The application of search engine optimization for internet marketing: An example of the motel websites. Proceeding of the IEEE 2nd International Conference on Computer and Automation Engineering (ICCAE), 1: 380-383.

Kavitha, R. and R. Nedunchelian, 2014. An efficient method to automatic intelligent based search engine in health care cloud data using c-means with neuro fuzzy system. Int. J. Appl. Eng. Res., 9(23): 22337-22350.

Killoran, J.B., 2013. How to use search engine optimization techniques to increase website visibility. IEEE T. Prof. Commun., 56(1): 50-66.

Li, K., M. Lin, Z. Lin and B. Xing, 2014. Running and chasing-the competition between paid search marketing and search engine optimization. Proceeding of the 47th Hawaii International Conference on System Sciences (HICSS), pp: 3110-3119.

Li, M., S. Yu, N. Cao and W. Lou, 2011. Authorized private keyword search over encrypted data in cloud computing. Proceeding of the 31st International Conference on Distributed Computing Systems (ICDCS), pp: 383-392.

Lim, E.H.Y., H.W.K. Tam, S.W.K. Wong, J.N.K. Liu and R.S.T. Lee, 2009. Collaborative content and user-based web ontology learning system. Proceeding of the IEEE International Conference on Fuzzy Systems (FUZZ-IEEE, 2009), pp: 1050-1055.

Lin, J. and S. Zhang, 2009. An optimizing search based on kernel-based fuzzy c-means clustering. Proceeding of the IEEE International Conference on Computational Intelligence and Software Engineering (CiSE, 2009), pp: 1-3.

Lin, T.F. and Y.P. Chi, 2014. Application of webpage optimization for clustering system on search engine v google study. Proceeding of the IEEE International Symposium on Computer, Consumer and Control (IS3C, 2014), pp: 698-701.

Liu, Y., Q.X. Wang, L. Guo, Q. Yao, N. Lv and Q. Wang, 2007. The optimization in news search engine using formal concept analysis. Proceeding of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD, 2007), 2: 45-49.

Majhi, S.K. and P. Bera, 2014. OHMF: A query based optimal healthcare medication framework. Int. J. Inform. Process., 8(3): 1-12.

Montesi, D. and A. Trombetta, 1999. Similarity search through fuzzy relational algebra. Proceeding of the 10th International Workshop on Database and Expert Systems Applications, pp: 235-239.

Nimbalkar, R.A. and R.A. Fadnavis, 2014. Domain specific search of nearest hospital and healthcare management system. Proceeding of the Recent Advances in Engineering and Computational Sciences (RAECS, 2014), pp: 1-5.

Sano, Y., H. Kita, I. Kamihira and M. Yamaguchi, 2000. Online optimization of an engine controller by means of a genetic algorithm using history of search. Proceeding of the 26th Annual Conference of the IEEE Industrial Electronics Society (IECON, 2000), 4: 2929-2934.

Santos, V., N. Datia and M.P.M. Pato, 2014. Ensemble feature ranking applied to medical data. Proc. Technol., 17: 223-230.

Shah, A. and S. Jain, 2011. An agent based personalized intelligent E-learning. Int. J. Comput. Appl., 20(3): 40-45.

Shi, L. and J. Sun, 2008. Automatic instance identification in intelligent information search engine CRAB. Proceeding of the IEEE 19th International Workshop on Database and Expert Systems Application (DEXA'08), pp: 175-179.

Shin, Y., J. Lim and J. Park, 2012. Joint optimization of index freshness and coverage in real-time search engines. IEEE T. Knowl. Data En., 24(12): 2203-2217.

Somani, A. and U. Suman, 2011. Counter measures against evolving search engine spamming techniques. Proceeding of the IEEE 3rd International Conference on Electronics Computer Technology (ICECT, 2011), 6: 214-217.

Sultana, S.N., G. Ramu and B. Eswara Reddy, 2014. Cloud-based development of smart and connected data in healthcare application. Int. J. Distrib. Parall. Syst., 5(6).

Tho, Q.T., S.C. Hui, A.C.M. Fong and T.H. Cao, 2006. Automatic fuzzy ontology generation for semantic web. IEEE T. Knowl. Data En., 18(6): 842-856.

Trappey, A.J.C., C.Y. Fan, C.V. Trappey, Y.L. Lin and C.Y. Wu, 2012. Intelligent recommendation methodology and system for patent search. Proceeding of the IEEE 16th International Conference on Computer Supported Cooperative Work in Design.

Venkatraman, S. and S.J. Kamatkar, 2013. Intelligent information retrieval and recommender system framework. Int. J. Future Comput. Commun., 2(2): 85-89.

Wang, F., L. Yi and Y. Zhang, 2011. An empirical study on the search engine optimization technique and its outcomes. Proceeding of the 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC, 2011), pp: 2767-2770.

Wang, R., Y. Chen, T. Li and Y. Yu, 2013. The optimization of search engines ranking technology based on grey system. Proceeding of the 5th International Conference on Computational and Information Sciences (ICCIS, 2013), pp: 1698-1700.

Yang, C.C., J. Yen and H. Chen, 2000. Intelligent internet searching agent based on hybrid simulated annealing. Decis. Support Syst., 28(3): 269-277.

Yang, K.H., C.C. Pan and T.L. Lee, 2003. Approximate search engine optimization for directory service. Proceeding of the International Parallel and Distributed Processing Symposium, pp: 8.

Yu, W.D. and A. Lin, 2007. The design and implementation of a Search Engine Marketing Management System (SEMMS) based on service-oriented architecture platform. Proceeding of the IEEE International Conference on e-Business Engineering (ICEBE, 2007), pp: 513-519.

Yunfeng, M., 2010. A study on tactics for corporate website development aiming at search engine optimization. Proceeding of the 2nd International Workshop on Education Technology and Computer Science (ETCS, 2010), 3: 673-675.

Zhang, Y., 2009. Result optimization returned by multiple Chinese search engines based on XML. Proceeding of the International Conference on Computational Intelligence and Software Engineering (CiSE, 2009), pp: 1-3.

Zhu, C. and G. Wu, 2011. Research and analysis of search engine optimization factors based on reverse engineeing. Proceeding of the 3rd International Conference on Multimedia Information Networking and Security (MINES, 2011), pp: 225-228.