Modified Firefly Algorithm and Fuzzy C-Mean Clustering Based Semantic Information Retrieval

M. Subramaniam*, A. Kathirvel, E. Sabitha† and H. Anwar Basha

Department of CSE, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Vadapalani, Chennai, Tamilnadu, India
E-mail: subramam2@srmist.edu.in; kathirva@srmist.edu.in; se7648@srmist.edu.in; anwarbah@srmist.edu.in
**Corresponding Author; †Research Scholar*

Received 29 September 2020; Accepted 31 October 2020;
Publication 15 February 2021

Abstract

As enormous volume of electronic data increased gradually, searching as well as retrieving essential info from the internet is extremely difficult task. Normally, the Information Retrieval (IR) systems present info dependent upon the user's query keywords. At present, it is insufficient as large volume of online data and it contains less precision as the system takes syntactic level search into consideration. Furthermore, numerous previous search engines utilize a variety of techniques for semantic based document extraction and the relevancy between the documents has been measured using page ranking methods. On the other hand, it contains certain problems with searching time. With the intention of enhancing the query searching time, the research system implemented a Modified Firefly Algorithm (MFA) adapted with Intelligent Ontology and Latent Dirichlet Allocation based Information Retrieval (IOLDAIR) model. In this recommended methodology, the set of web documents, Face book comments and tweets are taken as dataset. By means of utilizing Tokenization process, the dataset pre-processing is carried

Journal of Web Engineering, Vol. 20_1, 33–52.
doi: 10.13052/jwe1540-9589.2012

out. Strong ontology is built dependent upon a lot of info collected by means of referring via diverse websites. Find out the keywords as well as carry out semantic analysis with user query by utilizing ontology matching by means of jaccard similarity. The feature extraction is carried out dependent upon the semantic analysis. After that, by means of Modified Firefly Algorithm (MFA), the ideal features are chosen. With the help of Fuzzy C-Mean (FCM) clustering, the appropriate documents are grouped and rank them. At last by using IOLDAIR model, the appropriate information's are extracted. The major benefit of the research technique is the raise in relevancy, capability of dealing with big data as well as fast retrieval. The experimentation outcomes prove that the presented method attains improved performance when matched up with the previous system.

Keywords: Ontology, semantic information, web documents and modified firefly algorithm.

1 Introduction

Huge volume of documents has been indexed in the web which can be accessed by various users. The size of document set has been growing every day. As a result, the capability of accessing as well as choosing appropriate info in this enormous as well as heterogeneous volume of data continues to be a complex task [1]. On the other hand, a lot of Information retrieval systems contain some degree of abilities to use the conceptualizations in user requirements as well as content meanings [2]. It has some disadvantages for instance the lack of ability to define associations amid search terms. With the intention of overcoming these drawbacks, the modern researchers focused on extracting information in efficient way [3].

The foremost conception is towards extracting the content according to the conceptual similarities. Two foremost types are there in concept based document retrieval. The primary one consider the semantic relations between the query and documents by means of examining the latent associations amid text words. The next type encompasses techniques, which physically build taxonomy of semantic conceptions and their association with the query. In the second type, Ontology is one among well-known technologies also called as a knowledge representation. The impact of ontology in the information retrieval has monitored high and has to be considered as important factor. However, the application of ontology can be minimized [4]. Also, it can be

used in both the phases of information retrieval by presenting the knowledge as ontology [5]. They are very proficient particularly by means of utilizing domain-information extraction. On the other hand, they utilize different representation and language to perform document retrieval which is highly complicated. Formulating a query with the help of such languages needs the acquaintance of the domain ontology in addition to the syntax of the language.

In general the document retrieval algorithms uses the keywords to perform matching towards the terms of any web document. The method would select set of web pages which are approximately close to the input query term. Web provides certain chance to get better the conventional search. A better means is by using semantic search. The application of semantic search increases the accuracy in a huge space of data. The main objective of the semantic based retrieval system is to produce result based on the interest of the user with higher relevancy.

Semantic Information Retrieval is used to a lot of techniques dependent upon their characteristics and features [6]. The context based approaches uses semantic information to perform clustering in hybrid manner [7]. The profile based approach uses the semantic information to identify the interests related to the user which has been used to produce personalized result to the user. Analogously to researches comprising implicit feedback techniques in IR that have identified that profiles dependent upon the content of clicked URLs outdoes those dependent upon previous queries along, it identified that profiles dependent upon the content of bookmarked URLs are better than those dependent upon tags alone [8]. In Query based method, Semantic Information Retrieval is designed to provide user personalization in web search to improve the performance in recommendation generation [9]. It presents a ranking algorithm which ranks the pages according to the visit frequency of various users. The system would log the user visits and based on that the pages can be ranked [10, 11].

Numerous systems implemented a Semantic Similarity Retrieval Model (SSRM) with the aim of attaining greater relevancy. SSRM recommends finding semantically identical terms in documents utilizing various domain ontology which can be classified based on generic category or based on the application. The method initially estimates the frequency of terms in the document and the frequency of term in other class documents to compute Tf and IDF values. Finally a weight measure has been computed to classify the query class to perform document retrieval.

The information content techniques are modeled with the intention of getting high semantic similarity measurement [12]. In this proposed method, the similarity has been measured based on the value of MeSh.

2 Literature Survey

Pablo Castells et al. presented a document retrieval algorithm which uses the semantic ontology towards the user personalization in document search. The documents features are converted into vector forms and cosine similarity has been measured towards different class of documents. Based on the cosine similarity, a semantic weight has been measured to perform selection of class. According to the selected class, a set of result has been returned [13].

Sa and Varghese considered that it is not necessary that the document of the class should speak about the category in complete manner. The fact of topical discussion would vary and they would fall in a range. Towards this, a fuzzy model for document retrieval with semantic ontology has been discussed [14].

Sonar et al. discusses a document retrieval approach which takes a keyword as input to produce set of results. The documents are classified under several classes based on their semantic concept discussed. Based on the semantic classes, an query relevancy has been measured and set of results has been produced [15].

Fouad et al. presented personalized system for web document retrieval (SPIRS) which works based on the agents meant to perform document retrieval. The agents estimates the frequency of concept being discussed on the document and estimates weights for each document. According to the weight estimated a subset of documents has been returned as result [16]. Similarly Luo and Xue [17] incorporated multiple agents in the process of document retrieval. The method uses both conceptual features, user behaviors and the feedbacks obtained from various users in generating results to the user.

Hu et al. [18], presented a rough set theory based semantic document retrieval algorithm which consider the semantic relations between the search keyword and the documents of the class.

Zidi and Abed presented a rule based approach with semantic index for efficient document retrieval problem. The method indexes the documents in two categories like basic and rule. According to the index, an graph has been generated and towards the query, the method estimates graph feature similarity to identify related documents [19]. Similarly Fernandez et al.

presented an semantic based document retrieval system for unstructured documents in [20].

3 Proposed Methodology

The set of web documents, Face book comments and tweets are taken for evaluation. An interface has been designed to collect web documents and the tweets. Find out the keywords and carry out semantic analysis with user query utilizing ontology matching by means of jaccard similarity. Dependent upon the similarity the ideal features are chosen with the help of Modified Firefly Algorithm (MFA). After that, by utilizing fuzzy c-mean clustering, the appropriate documents are grouped. At last, use an Intelligent LDA to extract the appropriate documents. An Intelligent LDA based retrieval algorithm combines three other routines like clustering, page ranking and Intelligent LDA. Clustering algorithm creates various class of documents dependent upon the similarity and ranking of the document towards the class. At last, the intelligent LDA examine as well as take out the appropriate document content, which are dependent upon the queries. The research model denotes the knowledge base dependent upon user query. In Figure 1, architecture of the presented technique for the system designed for social networks analysis.

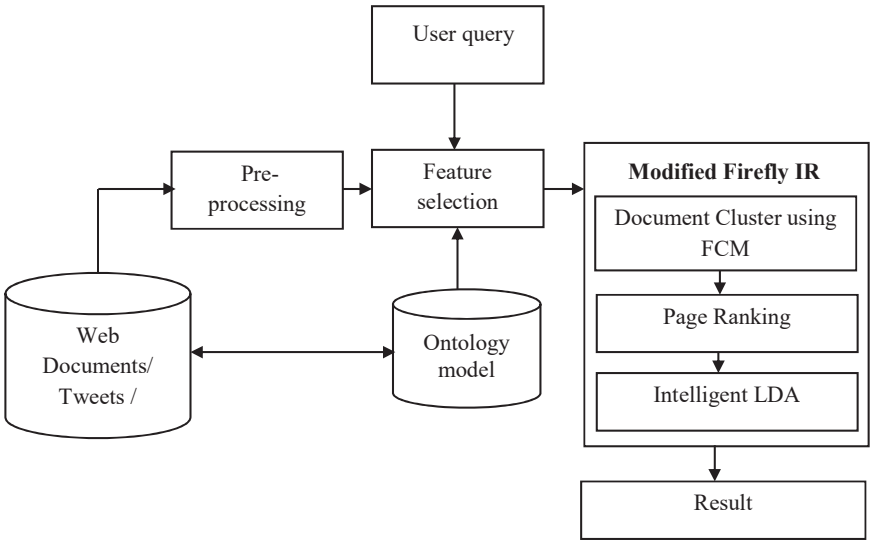

Figure 1 General Flow Diagram of modified firefly algorithm based IR.

3.1 Query Pre-processing

In this stage, the input document has been read and the textual features has been extracted. The text features are tokenized to produce a term set. From the term set, the method removes the list of terms which has no meaning like stop words. Then by applying the tagging process, the terms which are not a root word has been removed. This reduces the search time complexity and reduces the search space. In this research work, a method is presented for of tokenization, wherein token identification is entirely dependent upon the documents vectors.

The result of tokenization has been used to estimate the term count. The estimated count has been used to perform clustering or indexing the document towards a class [7, 8]. In conjunction with token generation this process computes the frequency measure of all the terms in the document given. In Figure 2, the phases of tokenization process are depicted. Pre-processing comprises the collection of all document and used to perform term extraction [12]. In subsequent phase all the infrequent words are queued as well as eliminated, such as elimination of word containing frequency below two. Intermediate outcomes are input to the subsequent step that is stop word elimination phase.

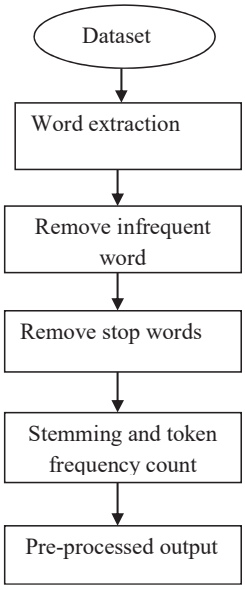

Figure 2 Preprocessing process.

The Figure 2, shows the stages of preprocessing algorithm which removes the stop words from the term set generated. Then for each term present in the term set, the method applies a stemming process. The result of stemming is used to identify the root words or pure nouns. Finally, the frequency measure has been computed. The preprocessed output has been used to perform clustering.

3.2 Ontology Creation and Semantic Matching

The semantic ontology represent the concepts of various domains which incorporates the concept and relations between them. It offers a general overview of a concept and as well its association with other concepts or terms. Therefore a hierarchy is created with the associated terms. The semantic web utilizes ontology as a tool to take concepts for particular domains. It is a working model of entities as well as interactions generally or in some specific domain of knowledge or practice. Ontology concepts stand for a collection or class of entities or things within a domain. It encompasses finite set of term as well as association amid these terms. For the web, ontology is precise depiction of web info or social media info and associations amid web or social media info. In this proposed method, find out the keywords and carry out semantic analysis amid query words and words in the ontology utilizing Jacard similarity measure.

Jacard similarity: The Jaccard coefficient is measured based on the frequency of terms being present and the frequency of terms being shared between different categories [14].

$$\text{SIMj} = (\vec{t_a}, \vec{t_b}) = \frac{(\vec{t_a}, \vec{t_b})}{(|\vec{t_a}| + |\vec{t_b}|) - (|\vec{t_a}| \cdot |\vec{t_b}|)} \tag{1}$$

The value of the coefficient is between 0–1.

3.3 Modified Firefly Algorithm

Dependent upon the semantic analysis, the features are taken out and the significant features are chosen by means of utilizing Modified Firefly algorithm. The method works based on the inspiration of biology can be used for optimization problem. Also it works according to the conduct of flash in the night. At each flash it identifies a superior one using an objective function. Similarly the method estimates the attractive of flash in each category.

The light intensity varies according to the distance. The coefficient of light absorption is λ:

$$I(r) = I_0 e^{-\lambda r} \tag{2}$$

Here, I_0 is source intensity and has been framed as

$$I(r) = I_0 e^{-\lambda r^2} \tag{3}$$

On the other hand, computationally computing $1/(1 + \lambda r^2)$ is simpler compared to computing $e^{-\lambda r^2}$. For this reason, the intensity is computed by means of utilizing

$$I(r) = \frac{I_0}{1 + \lambda r^2} \tag{4}$$

Likewise, the attractiveness of a firefly is defined in this manner:

$$A(r) = \frac{A_0}{1 + \lambda r^2} \tag{5}$$

Here, A_0 is the initial attractive.

Initially $x' = (x'_1, x'_2, \ldots, x'_n)$ is closure than $x = (x_1, x_2, \ldots, x_n)$, the firefly situated at x would move in the direction of x'. The updating of location is estimated as follows:

$$x := x + A_0\, e^{-\lambda r^2}(x' - x) + \alpha\varepsilon \tag{6}$$

The fine solution is the best firefly. To identify the most relevancy parameter, the algorithm generates a random vector $u_1, u_2, \ldots, u_m$. Second a direction has been selected to move the firefly. The movement of firefly is presented as follows:

$$x := x + \alpha U \tag{7}$$

Here, α is the value of single movement.

If there is no direction found it produces a solution and continue in the same position.

Moreover, more willingly than considering $A_0^i = 1$ in each iteration it estimates the attraction value based on the intensity. However A_0 is provided by

$$A_0 = \frac{I'_0}{I_0} \tag{8}$$

Here I'_0 is value of intensity in the initial location. Practically, for removing the singularity case while $I_0 = 0$, A_0 is as well be defined as $e^{I'_0 - I_0}$.

When we consider $A_0 = I_0'$ and at higher intensity condition there will be higher firefly movement. On the other hand based upon the solution space it is superior to regulate A_0. In either case it must be directly relative to the intensity at the source, I_0'.

Firefly algorithm

1. Create starting population (number of features) x_i (1 to n)
2. Calculate documents topical strength (accuracy) of fireflies
3. Describe document concept intake coefficient γ
4. Till(t>Maximum Generation)
5. For i = 1 to N number of fireflies of features
6. For j = 1 to N number of fireflies of features
7. Document topical strength
8. if topical strength of j > topical strength of i
9. move the features of I towards j
10. Else
11. Move the feature of I arbitrarily
12. End
13. Attractiveness modifies with distance r via exp [$-\gamma$r2]
14. Identify novel solutions and revise topical strength
15. End End
16. Rank the documents and choose the best
18. end
19. Return the relevant features to the classifier.

Dependent upon the semantic relevancy the finest features are chosen. The chosen features are input to the clustering process.

3.4 Fuzzy C-Means Clustering

Subsequent to the completion of the finest feature selection, the appropriate documents are grouped with the help of FCM clustering technique. The scope is to identify center point of the cluster (finest value of the feature), which decrease a dissimilarity function. The matrix of membership (U) is has been initialized as follows:

$$\sum_{i=1}^{c} u_{ij} = 1, \quad \forall\, j = 1, \ldots, n \tag{9}$$

The similarity function utilized in FCM is provided in Equation (9)

$$J(U, c_i, i = 1, \ldots, c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \tag{10}$$

Here u_{ij} is amid 0 and 1; c_i is known as cluster center I (finest value of the feature); d_{ij} is known as the distance amid i^{th} centroid (c_i) and j^{th} data point (features); $m \in [1, 8]$ is known as a weight expo.

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{11}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \tag{12}$$

This algorithm identifies the subsequent steps.

1: Arbitrarily initialize the membership matrix (U), which contains restraints in Equation (9).
2: Compute finest value of the feature
3: compute irrelevancy against various centroids (10). Stop if enhancement $<$ threshold.
4: Calculate a new U. Go to Step 2.

By means of iteratively bringing up to date the cluster centers (finest value of the feature) and the membership grades for every data point (features), FCM iteratively moves the cluster centers to the "right" place within a data set.

With the help of FCM algorithm the appropriate documents are grouped dependent upon their semantic relevance. The major benefits of the research techniques are the increase in accurateness of the system.

3.5 Ranking and Information Extraction

The presented algorithm works as per LDA (Xing Wei & Bruce Croft 2006). The recently presented one is a document based that utilized to look for titles is classified based on the words and how they distributed. The samples are obtained from unique terms. The weight assignment is performed according to query terms and multinomial parameters are random variables φ and $\theta^{(d)}$ with conjugate Dirichlet priors.

Input: Input Query/ search term.

Output: Related data

1: Identify documents according to query/ontology similarity
2: cluster the documents based on the similarity
3: Rank documents according to the closure
4: Create Documents dependent upon a Dirichlet calculated by utilizing the values of β, ϑ_z created over z;
5: Towards each documents $doc(i) = \{D1, D2 \ldots, Dn\}$, carry out the subsequent steps:
5.a validate each class/sub class on all parameters
5.b If the document is closure as medium and high, perform additional analysis
(i) Estimate Dirichlet score
(ii) For each doc
(iii) Data from $\theta^{(doc)}$ for the title z_n;
(iv) Data from φ_{z_n} for the word w_n;
Else
Show message in it ontology traversal
8: End

Every text from document is close to the latent title z. The $\theta^{(d)}$ on z is found based on α and the title upon h.

4 Results and Discussion

The method has been evaluated for its dynamic document retrieval and compared with previous approaches. The method has been implemented using Advanced java. The method has been validated for its relevancy score in document retrieval towards varying number of groups and varying number of documents in each group. The range of number of documents in each group has been varied in thousands which start from one thousand to five thousand. Each group has been considered with number of tweets obtained from the twitter data set.

The comparison result on relevancy towards the document retrieval has been measured and presented in Table 1. The proposed approach has improved the accuracy in relevant result higher than other methods. It has produced efficient result on all the number of document class and tweets considered.

Table 1 Comparison on relevancy

Test Case	Classes	Size of Document Set	Size of Tweet Set	Relevancy Measure (%)	
				Web Pages.	Tweets
1	Class1	1500	1500	97. 43	99.51
2	Class2	3000	3000	97.68	99.41
3	Class3	5000	5000	98.23	99.70
4	Class4	7500	7500	98.62	99.20
5	Class5	10000	10000	98.95	98.30

Table 2 Analysis on time complexity with web pages

No. of Records Considered	Latent Dirichlet		Semantic Orient Retrieval		IOLDA–IR		MFA with IOLDA–IR	
	Entire Features	Selective Features	Entire Features	Selective Features	Entire Features	Selective Features	Entire Features	Selective Features
2000	0.31	0.29	0.28	0.23	0.24	0.19	0.21	0.18
4000	0.43	0.35	0.35	0.28	0.26	0.22	0.23	0.21
6000	0.54	0.38	0.39	0.32	0.29	0.27	0.26	0.23
8000	0.58	0.42	0.43	0.35	0.34	0.29	0.28	0.24
10000	0.63	0.46	0.48	0.39	0.38	0.34	0.30	0.28

Table 3 Comparison on feature selection with tweets

No. of Records Considered	Latent Dirichlet		Semantic Orient Retrieval		IOLDA–IR		MFA with IOLDA–IR	
	Entire Features	Selective Features	Entire Features	Selective Features	Entire Features	Selective Features	Entire Features	Selective Features
2000	88.92	92.14	89.03	92.43	92.43	99.21	99. 37	99.39
4000	89.54	91.27	90.07	92.38	92.36	99.31	99. 38	99.41
6000	89.57	91.33	90.13	92.38	92.38	99.41	99.43	99.45
8000	89.63	91.35	90.16	92.43	92.43	99.45	99.45	99.46
10000	89.66	91.37	90.19	92.46	92.46	99.47	99.48	99.51

The performance on time complexity has been measured and compared with different approaches. The result of time complexity analysis is presented in Table 2.

According to Tables 3 and 4, it is clear that the presented classifier functions superior to the previous classification algorithms on selective and entire features.

Table 4 Analysis on feature selection with Documents

No. of Records Considered	Latent Dirichlet		Semantic Orient Retrieval		IOLDA–IR		MFA with IOLDA–IR	
	Selective Feature	Grouped Feature	Selective Feature	Grouped Feature	Selective Feature	Grouped Feature	Selective Feature	Grouped Feature
1000	91.24	91.26	92.34	92.36	99.12	99.14	99.16	99.18
2000	91.26	91.28	92.37	92.39	99.15	99.16	99.19	99.21
3000	91.31	91.33	92.39	92.41	99.18	99.18	99.26	99.27
4000	91.33	91.35	92.42	92.44	99.21	99.22	99.30	99.32
5000	91.36	91.38	92.45	92.47	99.24	99.25	99.35	99.37

4.1 Relevancy Score Analysis (Web Documents)

Relevance states how well a retrieved document or collection of documents meets the information requirement of the user. Semantic information retrieval technique contains the benefits of the semantic web for retrieving the appropriate data.

Relevancy Analysis

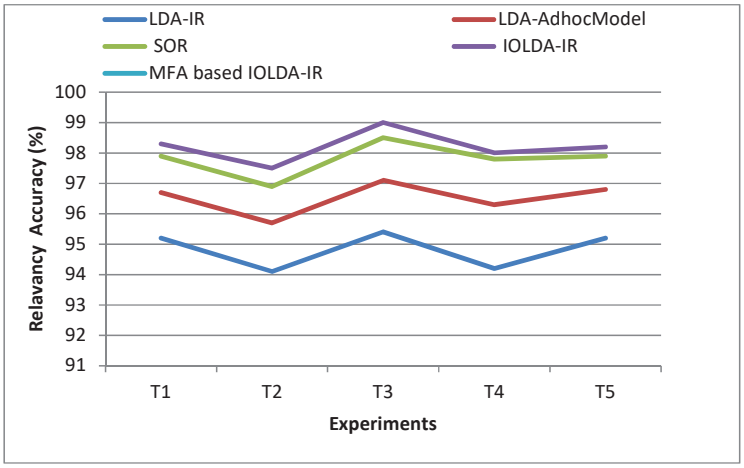

Figure 3 Comparison on relevancy score (web documents).

Figure 3 specifies that the pictorial depiction of Relevancy accuracy for previous and presented techniques. In x-axis a variety of experimentations

are considered and in y-axis Relevancy Score for Web Documents are considered. In this research, the MFA algorithm is used to choose the ideal features for semantic information retrieval process. And the fuzzy c-means algorithm is utilized for grouping the appropriate documents as per the semantic meanings. The experimentation outcomes prove that the presented method attains greater accuracy for semantic information mining in web documents.

4.2 Relevancy Score Analysis (Tweets)

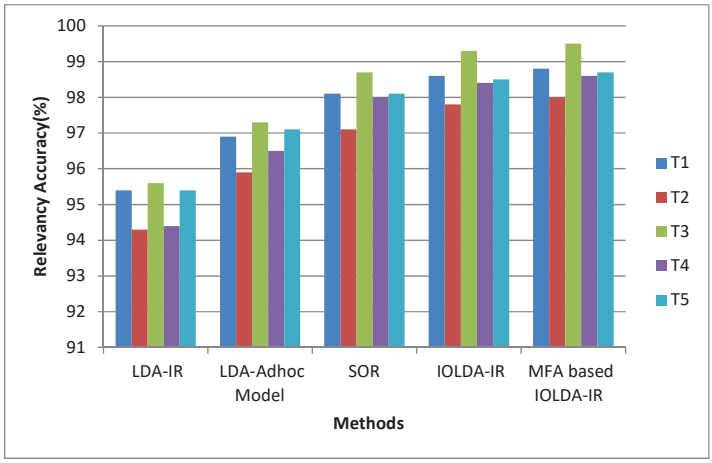

Figure 4 Relevancy score analysis (tweets).

Figure 4 depicts the relative analysis in regard to relevancy accuracy of the presented MFA based IOLDA-IR model and the previous LDA-IR, LDA adhoc model, SOR and IOLDA-IR on Tweets. Numerous experiments are considered in x-axis and in y-axis Relevancy Score for Tweets are considered. In this research, the relevant documents are grouped with the help of FCM clustering technique to get more appropriate outcomes. It enhances the relevancy score of the system. The experimentation outcomes prove that the presented method attains greater accuracy for semantic information mining in web documents.

4.3 Relevancy Score Analysis (Face book Messages)

According to Figure 5, it is clear that this presented MFA based IOLDA-IR model works well in a variety of experiments with diverse amount of Face

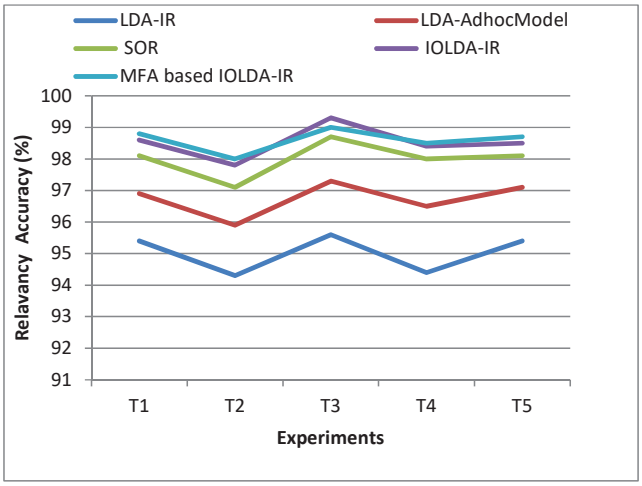

Figure 5 Relevancy score analysis (face book messages).

book messages while it is matched up with other previous LDA-IR, LDA adhoc model, SOR and IOLDA-IR models. A variety of experiments are considered in x-axis and in y-axis Relevancy Score for face book messages are considered. In this research, find out the keywords and carry out semantic analysis with user query utilizing ontology matching by means of jaccard similarity. Dependent upon the semantic analysis the feature extraction is carried out to enhance the on the whole retrieving performance. The experimentation outcomes prove that the research method attains greater accuracy for semantic information mining in web documents. This is owing to the reality that the usage of ontology, temporal features and MFA based feature selection.

5 Conclusion

The problem of semantic search has been a challenging investigation being considered in the modern research. The goal of the presented semantic retrieval model is conversed to offer improved search abilities, which will provide a qualitative enhancement over keyword-based complete text search, by presenting Modified Firefly Algorithm based feature selection, fine-grained domain ontologies with keyword matching. The structural design of the presented method is flexible to handle with user defined ontology which handles the generic query using standard query languages. The

experimentation outcomes prove that the presented system attains improved performance when matched up with the previous methodology.

Conflicts of Interest: None

References

[1] Marie-Aude, Aufaure, On-Line semantic Infolrmation retrieval using Ontologies, IEEE, 2007.
[2] Zheng, and Song-Nian, Ontology-Based Inverted Tables in Information Retrieval System, ICSKG, pp. 354–357, 2007.
[3] Been-Chian Chien, Chih-Hung Hu Intelligent Information Retrieval Applying Automatic Constructed Fuzzy Ontology, ICMLC, pp. 2239–2244, 2007.
[4] C. Carpineto, G. Romano. A Survey of Automatic Query Expansion in Information Retrieval. ACM Comput. Surv. 44, 1, 2012.
[5] K. Soner K., A. Özgür, An ontology-based retrieval system using semantic indexing. Inf. Syst. Vol. 37, issue 4, pp 294–305, 2004.
[6] Castells, P., Fernandez, M. An adaptation of the vector-space model for ontology-based information retrieval, *IEEE knowledge and data engineering*, vol. 19, iss. 2, 2007.
[7] Varelas, G., Voutsakis, E., Semantic similarity methods in wordNet and their application to information retrieval on the web. *ACM international workshop on Web information and data management* (pp. 10–16), 2007.
[8] Shyu, C. R., Klaric, GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, and complex queries. *IEEE Transactions on geoscience and remote sensing*, *45*(4), 839–852, 2007.
[9] Zhuhadar, Leyla, Olfa Nasraoui. "Semantic information retrieval for personalized e-learning." IEEE Conference on *Tools with Artificial Intelligence*, 2008.
[10] Rinaldi, A. M. An ontology-driven approach for semantic information retrieval on the web. *ACM Internet Technology (TOIT)*, vol. 9, iss. 3, 2009.
[11] Kumar, M.S., N. Prakash Developing university an ontology in education domain using protégé for semantic web. Int J. Eng. Sci. Technol., 2: 4673–4681, 2010.

[12] Lord, P., Stevens, R., Investigating Semantic Similarity Measures across the Gene Ontology: the Relationship between Sequence and Annotation. Bioinformatics, vol. 19, iss. 10, pp. 1275–83, 2003.

[13] Pablo Castells, Miriam Fernańdez, Self-tuning Personalized Information Retrieval in an Ontology-Based Framework, International conference on "On the Move to Meaningful Internet Systems", paper 11.3.4, pp. 977–986, 2005.

[14] Remi, S., & Varghese, S. C. Domain ontology driven fuzzy semantic information retrieval. Procedia Computer Science, 46, 676–681, 2015.

[15] Kara, Soner, et al. An ontology-based retrieval system using semantic indexing, Information Systems, vol. 37, iss. 4, pp. 294–305, 2012.

[16] Fouad, K. M., Khalifa, Web-based Semantic and Personalized Information Retrieval Semantic and Personalized Information Retrieval Semantic and Personalized Information Retrieval, 2012.

[17] J. Luo, X. Xue. Research on Information Retrieval System Based on Semantic Web and Multi-Agent, International Conference on Intelligent Computing and Cognitive Informatics, 2012.

[18] Hu, J., Lu, X., & Guan, C. A Semantic Information Retrieval Approach Based on Rough Ontology. *The Open Cybernetics & Systemics Journal*, vol. 8, pp. 399–404, 2014.

[19] Zidi, A., & Abed, M. A generalized framework for ontology-based information retrieval: Application to a public-transportation system. In *Advanced Logistics and Transport (ICALT)*, 2013.

[20] Fernández, M., Cantador, I., López, V., Vallet, D., Castells, P., & Motta, E. Semantically enhanced information retrieval: An ontology-based approach. Web semantics: Science, services and agents on the world wide web, 9(4), 434–452, 2011.

Biographies

M. Subramaniam (1974) is a Professor, in Department of Computer Science and Engineering, School of Computing, SRM Institute of Science and Technology (Deemed to be University u/s 3 of UGC Act, 1956)- Vadapalani Campus, Chennai- 600026, (INDIA). He obtained his Bachelor's degree (B.E) in Computer Science and Engineering from University of Madras (1998), Master degree (M.E) in Software Engineering and Ph.D from College of Engineering-Guindy (CEG), Anna University Main Campus, Chennai -25 in the year 2003 and 2013 respectively. His research focuses are Computer Networks, Software Engineering, AI& ML. He is an active life member of the Computer Society of India (CSI), the Indian Society for Technical Education (ISTE) and International Association of Engineers (IAENG). He has produced one doctorate and currently seven research scholars pursuing Ph.D under his guidance. He has published many research papers in reputed journals. He is also reviewer in Springer- WPC, IEEE- International Journal of Communication Systems.

Kathirvel Ayyaswamy, acquired, B.E.(CSE), M.E. (CSE) from University of Madras and Ph. D (CSE.) from Anna University. He has served in various positions at Deemed Universities, Autonomous Institution and Anna University affiliated colleges from 1998 to till date. He is currently working

as Professor, Dept of Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani Campus at Chennai. He has worked as Lecturer, Senior Lecturer, Assistant Professor, Professor, and Professor & Head in various institutions. He is a studious researcher by himself, completed 18 sponsored research projects worth of Rs.103 lakhs and published more than 110 articles in journals and conferences. 4 research scholars have completed Ph. D and 3 under progress under his guidance. He is working as scientific and editorial board member of many journals. He has reviewed dozens of papers in many journals. He has author of 12 books. His research interests are protocol development for wireless ad hoc networks, security in ad hoc network, data communication and networks, mobile computing, wireless networks and Delay tolerant networks.

E. Sabitha, Research Scholar, is pursuing her research in the Department of Computer Science & Engineering, SRM Institute of Science & Technology, (Deemed to be University u/s 3 of UGC Act, 1956) – Vadapalani Campus, Chennai-600026, (INDIA). She obtained her B.Tech(IT) from Anna University (2011), Chennai. She has obtained her M.E (CSE) from St.Peter's University (2013), Chennai. Her area of research is Artificial Intelligence and Machine Learning. She has 5 years of teaching experience. She has pubished papers in various National/International Conferences.

H. Anwar Basha is working as an Assistant Professor in the Department of Computer Science & Engineering, SRM Institute of Science & Technology, Vadapalani. He has obtained his B.E degree from Anna University, Chennai. He has obtained his M.Tech degree from Dr. MGR Educational and Research Institute University, Chennai. He has more than 12 + years of teaching experience. He has around 2 Years of Industrial Work experience. He has published papers in various International conferences and peer-reviewed international journals. Currently, He is working on Multi-Cloud Storage and Cyber Security.