Received: January 5, 2020. Revised: February 4, 2020.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

An Ontology Based Framework for Automatic Web Resources Identification

Eman El-Sayed Mahmoud[1]* **Ashraf Soliman Mohamed[2]**

[1]*Modern University for Technology and Information, Egypt*
[2]*October University for Modern Sciences and Arts, Egypt*
* Corresponding author's Email: emanelsayedmahmoud@gmail.com

Abstract: Search engines' capabilities are improved continually for enhancing the retrieved results and satisfying users' needs. Despite the great efforts in the information retrieval field, the user spends long time in changing the search keywords; though the retrieved results may be out of user's expectation. This may be due to the huge number of web resources, the lack of search keywords for specific domains, and the gap between the user and author in specific domain. This paper proposes a model for improving the search engines capabilities through adding level of awareness to the web resources. The model targets the human view about web resources in addition to author perception. This is through automatically extraction of the user's perceptions and mapping them to the resources' based on semantic aspects. The experimental studies realized 36% precision improvement compared to other search engines, and user satisfaction was up to 92%.

Keywords: Topic ontology, Web resources, Social tags.

1. Introduction

Recently, Information Retrieval (IR) researchers' efforts are directed to enhance the retrieval models for the purpose of satisfying the user's needs rather than web users spend a long time modifying their search keywords in order to reach their desired results. This challenge comes from the search engines matching techniques, where the most popular search techniques are depended on keyword matching [1, 2, and 3]. Furthermore, the most existing sensitive information retrieval models focus on improving the retrieval decision through queries, resource keywords, implicit feedback, and users' clicks [2, 14]. Practically, investigation has indicated its poor user experience on Google search for 52% of 20,000 queries; searches did not find any relevant results [4].

In the context of Web 2.0, "users became part of the web not only recipients and several applications support that" [5, 9]. Nowadays, many platforms provide the ability to remark web resources by writing tags and annotations. Users' remarks may reflect additional awareness of web resources. Also, they add common expressions and abbreviations for specific field that play important role in enhancing retrieval.

This research area needs more investigation for supporting the topic based detection in specific domain. So, the proposed framework interests in enhancing the search engine capabilities by adding level of awareness of human speech and needs. Furthermore, the framework focuses on eliminating heterogeneity and negative navigation problem through developing a domain ontology. The developed ontology maps the web resources based on topics detection, semantic aspects, and users' annotations.

The proposed framework focuses on extracting the human speeches from the bookmarking systems that allow the web users to annotate web resources. It considers user's annotations as part of web resource index, then analyze them syntactically and semantically in order to map them to web resources automatically.

The human perception that shows in user's annotations and the author perspectives that reflected on the resource keyword; have a main roll in enhancing the search engine results and eliminating the gap between users and author.

The proposed framework and experiments are detailed in the following sections, where section 2 looks over the previous work in this area of research, section 3 touches on the proposed model in detail, and section 4 shows evaluation and experiments.

2. Related work

D. Yong. (2011) [8] proposed to exploit topic tag mining for enhancing information retrieval. Its improved language model based on three components; topic structures of documents, semantic structures of tags, and user interests. It calculated relation between three main parts social tags, resource, and web user. The result of calculation provided in mapping similar documents. Further, the resources similarity was calculated through, and clustering based on tag mining. So, it proposed to estimate the document model and rank results based on the query generative likelihood. It decomposed the model into four sub-models which combined together to develop query terms. The sub models were language annotation model, document model, user model, and query model. The proposed model realized improvement 10% compared to search engine results. Despite that it suffers from many challenges, where the user's special expression didn't concern as part of their scope. So, the gap between user and search engine keywords didn't fixed.

A. Rathore. (2014) [18] proposed an approach for automatic topic identification of web pages that can provide better results. The topic of the web documents is identified through ontological approach. Keywords are extracted from the basic HTML tags and co-occurrence of words in the text instead of calculating the frequency of each term exits in a web page. Domain ontology is developed to map topics of the documents. The result could give benefit to the search engines for faster tagging of web pages. The average of precision and recall are 71.4% and 40.5%. On the other side, the proposed approach didn't concern with the user as a part of analysis phase. It improved the search results based on developing an ontology that represented the topic keywords in specific domain.

M. Bouadjenek, (2016) [20] proposed a framework for enhancing the information retrieval. The proposed framework exploits annotation as a part of resource analysis in addition to the resource's content. The resource's content and tag analysis were analyzed based on syntactic analysis. Then, it used the vector model for measuring the similarity. Furthermore, the proposed approach was implemented in free datasets from different bookmarking systems such Delicious, Fliker, and CiteULike. The precision average is in the range 80% to 90%. In addition, the percentage of improvement compared to the search engines is in between 12% up to 21%. On the side, the proposed framework lacks of the semantic analysis for both document and user's annotations. This causes the ambiguity problem that affects the retrieval process.

M. Rani. (2017) [19] proposed two topic modeling algorithms are explored for learning topic ontology. The objective is to determine the statistical relationship between document and terms to build a topic ontology and ontology graph with minimum human intervention. Experimental analysis on building a topic ontology and semantic retrieving corresponding topic ontology for the user's query demonstrating the effectiveness of the proposed approach. Despite that, the model didn't focus on the semantic relationships extraction. Also, the user was not part of the analysis phase.

K. Batista. (2018) [17] proposed an application of a new ontology-based methodology for the automatic topic detection without any previous information based on the use of hierarchical clustering algorithms and a multilingual knowledge base. The approach also includes lexical resources that allow us to enrich the semantics of the analyzed texts. The novelty of this approach consists of the dimensionality reduction of the terms present in the texts by using ontology and the introduction of a method for the creation of a term weight matrix for use in clustering algorithms. Although this approach may improve automatic topic detection in documents, it could not meet the user expectation in many cases. This is due to the gap between the user and the search engine, where the user was out of their scope.

3. Proposed model

Search engines improve their capabilities toward minimize search time and efforts. One of the main challenges is keyword mismatching, especially academic resources suffer from two main challenges heterogeneity and unpopular search keywords. First, academic fields may have expressions and abbreviations that represent different contribution in different field. In addition, web users may have their own common words that interpret complex expressions or abbreviate it to easily exchange. In

such cases, search engines cannot meet the users' expectations, where the search keywords did not match resources. In this context, the proposed model considers the web resource's words are not enough to reflect its context.

The users' annotations on web resources enrich the resources with additional level of knowledge that could not be demonstrated through the author perception. Especially academic resource, the author utilizes the formal and academic keywords, whereas the user query may include common society expressions. Thus, the search engines task became too hard, where they need to eliminate the gap between users and academic resource nature.

The Semantic Resource Representation (SRR) is a proposed model aims to avoid the gap between user and formality of web resources. It exploits users' annotations for creating new level of domain awareness. It considers the users' annotations as a real measure that reflects the resource context. Further, it doesn't concern with the resources' keywords but those matched syntactically or semantically to annotations.

The key aspects of the SRR model are: (1) the web resources context based on semantic extraction. (2) the ontology driven representation for specific domain. (3) the use of automatic expert system in retrieval process. The SRR consists of two phases: **Preprocessing** phase and **Automatic Expert Retrieval (AER)** phase. The first phase concerns with the 1st and 2nd aspects, where the 3rd aspect is realized through the second phase.

3.1 Preprocessing phase

The preprocessing phase aims to develop a topic-based ontology that considered as the semantic based model constructed based on linguistic and semantic tag analysis. The Preprocessing phase consists of two stages: semantic context extraction, and automatic development of ontology as shown in Fig. 1.

The user can mark a resource by writing one or more tags on -it based on his interests. Users' annotations are written from the perspective of their understanding of resources' contents. So, the annotations' terms may be resource's keywords, common expressions or abbreviations for specific field. The annotation consists of three main parts tag, tagger, and annotated resources. The SRR model focuses on the annotated resources and tags regardless of who the tagger is.

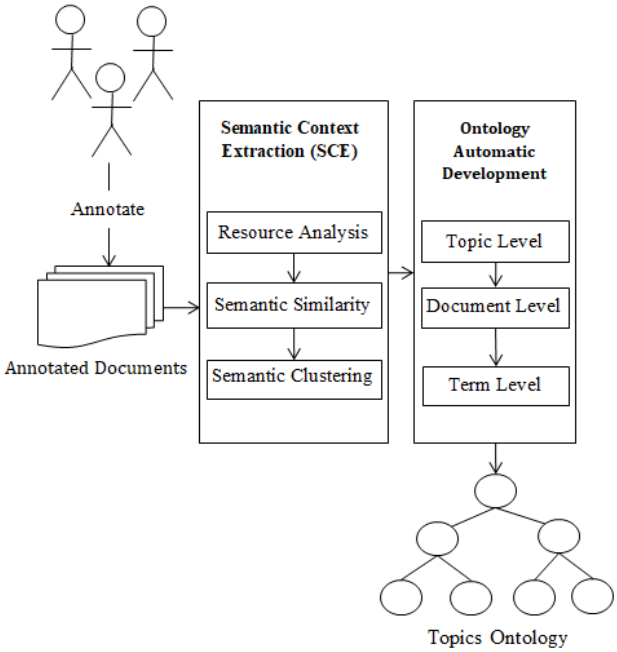

Figure. 1 The preprocessing phase of SSR model

Tag analysis is the prior step to reformulate the view of each resource. Linguistic and semantic techniques are exploited for analyzing tags. This provides more interpretation of tags and support semantic mapping.

3.1.1. Semantic context extraction (SCE)

The Semantic Context Extraction (SCE) is the basic step to reformulate the web resources' content view. Each resource content is represented through number of tags that have written by number of web users. In addition, the web resources' keywords are mapped to tags keyword depend on semantic sense. The SCE consists of three main steps: resource analysis, semantic similarity, and semantic clustering.

A. Resource analysis

First of all, resource analysis aims to collect and analyze all tags (T_1, T_2, T_i, …, T_n) that are attached to a resource (R_i) based on the classical information retrieval approach called *Term Frequency Inverse Document Frequency* (*TFIDF*) to weight tags' terms [16].

$$TFIDF\ (t,r,R) = TF\ (t,r)\ .IDF\ (t,R) \qquad (1)$$

By calculating the *Tag Term Frequency Inverse Document Frequency* (*TTFIDF*), an important challenge has to be solved, the non-meaningful terms that are included in user tags. Often, user may abbreviate common expression in specific filed and formally it's meaningless (e.g. text mining written

as txt mining). On the other hand, users may sign a resource by writing terms that do not relate to their domain at all. The challenge is how to distinguish valuable terms from the non-valuable ones.

The semantic based filtering process is applied for eliminating the non-valuable terms and focus on valuable - ones. The filtering process is linguistic filtering that exploits WordNet. The WordNet is a large online English lexicon for semantic checking tag's terms [16]. Many meaningful terms are added to resources' corpus, while useless terms are isolated and deleted such as misspelled words, and individual signs. In the context of determining the useful and useless terms the term's weight can be used. One of the challenges that face filtering is that the tagger may sign resource by useless term and repeat it in many tags. So, the repeated term has weight and may cause confuse. Thus, SRR determines a threshold of 0.5 to identify the useful terms that commonly used but has no meaning. Therefore, two types of terms will be added to resource corpus; the meaningful terms, and commonly used expressions that do not exist in WordNet but important for web users as search keywords.

B. Semantic similarity

Usually, tags suffer from heterogeneity and ambiguity problems, where taggers may use different terms in the same meaning or one term in different meanings. The Semantic Similarity measure integrates the syntactic and semantic features for improving the relationships between resources. Then, the resources are represented as vectors in the vector space model. In general, "the vector model suffers from some challenges like assumption of term indecency (e.g. ignore synonymy), and missing semantic information" [14]. Thus, the SRR model adds a semantic level to improve retrieval and overcome challenges of free text annotation.

In this context, the *Semantic Similarity Discovery (SSD)* algorithm concerns with extract the relationships between each pair of resources depended on the cosine similarity measure. SSD cosine similarity is a combination of syntactic and semantic similarity (Algorithm 1).

Algorithm 1. SSD algorithm

Input: terms' weights, and filtered tags
Output: cosine similarity based on syntax and semantic distance between resources
Process: extracting relevant terms and resources depended on WordNet semantic sense relations

FOR EACH annotated resource from R_i to R_n
  GET terms' weights for R_i, and R_j;
  WHILE (R_i & R_j still have terms);
    IF term T_i compared to T_j are equal **THEN**
      SR=1;
      Get (T_i weight) & (T_j weight *SR);
    ELSE IF T_j is synonym to T_i **THEN**
      SR=0.9;
      Get (T_i weight)&(T_j weight *SR);
    ELSE IF Tj is hyponymy to T_i **THEN**
      SR=0.5;
      Get (T_i weight)&(T_j weight *SR);
    ELSE
      Calculate cosine similarity for each pair of Resources R_i, R_j according to the modified weights;
    END IF
  END WHILE
END FOR

WordNet supports SSD by semantic sense for discovering synonym and hyponymy relations. So, the semantic sense leads to efficient improvement of resources similarity, where the resources similarity degree may become stronger or new resources relationships may be discovered. *SSD* adds the semantic sense to vector model by semantically comparing resources' terms, and categorizes them into four semantic relations as equal, semi-equal, partially equal, non-related. The equal relation is for terms that have the same syntax, semi-equal relation for synonyms, partially-equal relation for hyponymy, and non-related relation is the different terms. Each category has a pre-specified *Semantic Relation (SR)* degree which represents a threshold for identifying them. As in Eq. (1), the *Semantic Similarity Discovery (SSD)* identifies the web resources related degree using cosine similarity and *SR*.

$$SDD(R_j, R_i) = \frac{R_j \cdot Ri}{|R_j| \cdot |R_i|} = \frac{\sum_{i=1}^{t} (w_j * SR \cdot w_i)}{\sqrt{\sum_{i=1}^{t} (w_j^i * SR)^2 \cdot \sum_{i=1}^{t} w_i^2}} \quad (2)$$

The Euclidean lengths between web resources *(R_n)* is calculated by representing them as vectors through the vector space model. The *resources' term weights (w_n)* are calculated for each vector, and then the semantic relations are discovered. The *Semantic Relation (SR)* is added to vectors and mostly improves the similarity between resources.

C. Semantic Clustering

The purpose of this process is clustering documents based on topic. However, a document may be joined to multiple topics. Semantic Clustering is a basic step that leads to construct a resources profile. The development of resource profile depends on mining resources according to the semantic and syntactic analysis of tags. The SRR resource profile considers as combination of resources content and users' common knowledge.

The semantic clustering exploits the K-means algorithm and extends it. The additional two issues which added through the SRR are automatic seed identification, and semantic clustering. The automatic seed identification is automatically calculation of the k seeds by extracting the number of topics that completely different (as shown in Algorithm 2). Moreover, the semantic clustering integrates the topics which have semantic relationships (e.g. part of relation, synonyms). Finally, the output of the resource profile phase is represented in topic based ontology.

Algorithm 2. Clustering based on semantic sense

Input: resources R_n , semantic relationships
Output: resources clusters
Process: mining resources based on k-mean and additional semantic sense properties
Initialize core seed randomly;
WHILE (non-related seeds still exist)
  Get resources similarity;
  Compare initialized seed to all resources and their neighbors;
  Get all non-related points and check similarity with neighbors;
END WHILE
Identify array of k main point that completely different;
FOR i=1 To number of k
  FOR annotated resource from R_i to R_n
    Extract all neighbors to each k seed;
    Construct cluster;
  END FOR
END FOR
FOR all clusters
  GET intersected clusters;
  Construct intersected clusters' points;
END FOR

3.1.2. Topic-based ontology development

After Semantic Context Extraction, the resources are classified semantically based on their topic. Then, the outcomes form this phase are used to automatically constructing ontology. The Topic-Based Ontology (TBO) considers as a standard model that represent different resources in specific domain in consideration of their topics. The TBO structure as shown in Fig. 2 consists of three levels of knowledge; topics level, documents level, and terms level. This shows the semantic relationships among resources that may have intersected terms explicitly or implicitly. The explicitly relationships represents the context similarity. On the other side the implicitly relationships represent semantic relationships which discovered through the existence of synonym or hyponymy terms.

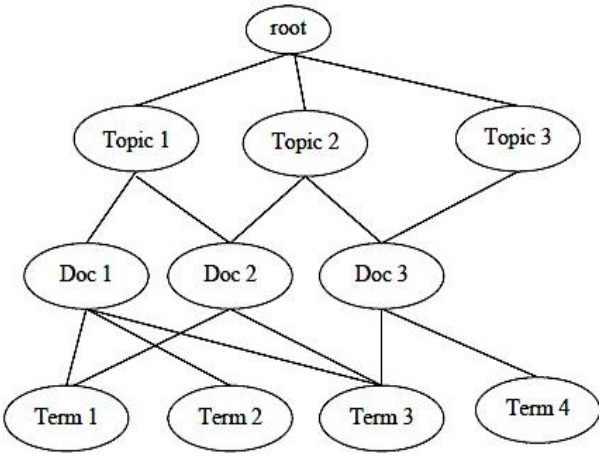

Figure. 2 Topic based ontology hierarchy

First of all, the ontology developed using top down technique where the outputs of semantic clusters are the main seeds to represent the top level. The TBO starts with topic seeds that represent different topics in specific field. Each topic is represented as a class and created automatically. Many resources are specified in one or more kind(s) of these topic but each one has different level of specialty. This measured through the vector model for determine the degree of specialty. The degree of specialty is important factor in ranking the user's query result (query answering is out of paper scope).

Secondly, the term level represents all indexed word that extracted from users' tags and the resources' related words. This level demonstrates the terms' weight based on semantic relationships. This level has two relationships layers; document to term (D2T) and term to term (T2T). D2T represents the relationship between a document and number of terms that mapped to document node based on syntactic or semantic relationships. Furthermore, the T2T represents the semantic relationships between terms.

4. Performance and evaluation

4.1 The SRR system performance

To evaluate the SRR model performance, java script and JAWS API are used. The SRR assumes the annotated resources have to be annotated at least by ten terms through more than three taggers, and ignore other cases. Further, it filters taggers for eliminating those whom have unclear interests.

The user's queries are applied through the SRR system which uses the topic-based ontology in two processes: query reformulation and ranking process.

Algorithm 3. Inferring similar terms

Input: user's query term (*tq)*
Output: number of related documents that semantically related to the query terms
Process: match the query terms semantically and extracting the related document
FOR EACH term t_i in the annotation ontology
IF equal (t_q, *ti*) **THEN**
    EqualSet ← t_i;
Else IF synonym (t_q, t_i) **THEN**
    SynonymSet ← t_i;
Else IF hyponym (t_q, t_i) **THEN**
    hyponymSet ← t_i;
ELSE
    nonRelatedSet ← t_i;
END IF
END FOR
FOR EACH term t_n in EqualSet
    Weight← getValue (t_n.*weights*);
END FOR
FOR EACH term t_n in SynonymSet
    Weight← getValue (t_n.*weights*);
END FOR
FOR EACH term t_n in HyponymSet
    Weight← getValue (t_n.*weights*);
END FOR

The query reformulation aims to expand the user's query based on syntax and semantic aspects. As shown in Algorithm 3, the user's query terms (*tq*) will be matched through the topic ontology using down-up technique. Then matched terms will be categorized into three sets (*EqualSet*, *SynonymSet*, and *hyponymSet).* First, the *EqualSet* includes the set of terms (t_i) that exactly matched the query term. Second, the *SynonymSet* includes the terms that matched the same meaning. Last, the *hyponymSet* holds the hyponym related terms.

The next step is identifying the related document using the above groups and calculating the weight of detected terms in each one. Based on that the documents will be ranked.

4.2 Evaluation

The model evaluation is done by applying the SRR system through CiteULIke dataset (http://www.springer.com/about+springer/ citeulike). CiteULIke bookmarking system allows users to tag several references (e.g., academic papers or books) included in its library. The CiteULIke dataset consists of documents directory that contains over 180 documents in text format, and annotated with 807 annotations.

The evaluation process assesses two factors: the user satisfaction and the retrieval improvement. First, the user satisfaction is one of the main SRR purposes, since users' expressions became part of the web resources. Thus, it is expected to retrieve results closer to users' expectations. Second, the retrieval improvement is measured by recall and precision.

The SRR model enhances query results by adding list of documents that could not retrieve through existing search engine matching techniques. By comparing the SRR model to search engine results, the SRR improves the resources' mapping similarity for 90% of resources and adds average of new discovered resources' relationships in percentage of 10%.

The SRR performance is measured using precision and recall. The target of SRR is to achieve high precision and low recall. The SRR approach has been applied into up to three hundred queries, and the number of query terms is between two to five terms. The query terms were tags' keywords, resources' keywords, and random expressions that are selected from the dataset. Practically, the SRR achieved a precision average in between 85.7% and 90.77%, and recall average in between 62.2% and 65.11%.

Further, for more evaluation, more than five hundred queries applied into both search engine system and the SRR system (as shown in Fig. 3. The search retrieved query results based on keyword retrieval technique, when the SRR integrates semantic and syntactic techniques. Then, the results are compared through precision and recall measures. The average of SRR precision and recall is 94.4% and 62.4%. Further, the SRR system achieved 85% of the all search engine results; so this approves that users' tags are considered a good reflection of resources. In addition, the average of precision improvement achieved through the SRR system compared to search engines is 36%.

Table 1. Comparison between SRR and previous work

	Ontology Development	Semantic Detection	User Participation	Precision Average	Improvement
SRR	Auto	√	√	94.6%	36.10%
Yong. [8]	NON	√	√	56.6%	10%
Rathore [18]	Auto	×	×	80%	32%
Bouadjenek [20]	Non	×	√	90%	21%
Rani [19]	Semi-Auto	×	×	75%	___
Basita [17]	Auto	×	×	30%	10%

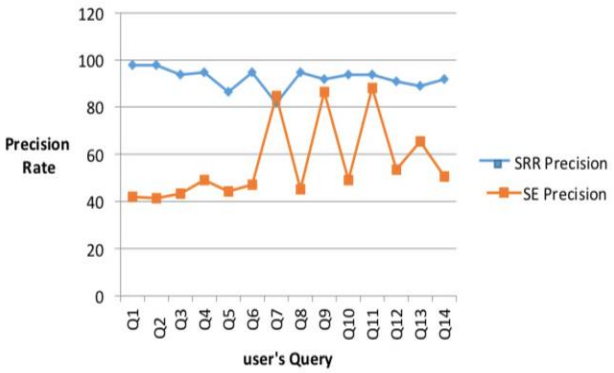

Figure. 3 The SRR & Search Engine precision

4.3 SRR versus previous work

In order to constructing comparison between the SRR approach and the previous work in the same research area; Table 1 shows comparison based on some criteria that represents the main features in the research field. First, the ontology based frameworks represent the level of ontology development in range of values between, automatic ontology development, semi-automatic development, and Non (not included ontology). Second, semantic detection criteria reflect the usage of semantic relationships and fixing the semantic problems. Third, the user participation criteria show the roll of user's perception in each model, where the user annotations or blogs consider as part of model construction and retrieval process. Last, the model efficiency is measured through precision average and rate of improvement compared to search engines retrieval process.

As shown in Table 1 the ontology development and topic detection for the SRR model, A. Rathore. (2014) [18], and K. Basita. (2018) [17] were done automatically except M. Rani. (2017) [19]. Further, the semantic relationships and fixing semantic problem (e.g. heterogeneity) are the main focus for the SRR and D. Yong. (2011) [8].

This in addition to consider the users' annotations and participation as part of web resources analysis in the SRR, M. Bouadjenek. (2016) [20], and D. Yong. (2011) [8]. Finally, the SRR model realized high/almost near precision average and improvement compared to search engines and the previous related work.

5. Conclusion

Nowadays, search engines direct their effort to enhance their capabilities by exploiting the web 2.0 benefits. The proposed model adds a level of human knowledge to the retrieval process by analyzing users' tags. The user's tags facilitate the resource's content discovering. Also, the user's tags reflect the domain knowledge, expressions, and uncommon search keywords. This contributes in retrieving results close to user's expectations.

Furthermore, the SRR adds levels of semantic and standardization by developing topic ontology. So, the SRR model improves the retrieval process based on two contributions: semantic sense, and topic identification. This improves user's query results and achieves high precision compared to other search engines in range (80%-94%).

The proposed future work will concern with integrating the intersected domains, where some sciences have shared knowledges and expression. This considers as challenge that cause ambiguity and affects the search engine results. In this context the future work aims to develop ontology alignment model for fixing the ambiguous expression between different domains; this may lead to improve the search results to be closer to the web user's expectations.

References

[1] G. Pasi, "Implicit Feedback Through User System Interactions For Defining User Models in Personalized Search", *Procedia Computer Science*, Vol. 39, pp. 8-11, 2014.

[2] M. Speretta and S. Gauch, "Personalizing Search Based on User Search Histories", *In: Pro. of International Conf. of Knowledge Management (CIKM),* Washington, USA, Vol. 2005, pp. 622-628, 2004.

[3] S. Waghmare and R. Krishna, "Implementation of Personalized Search Model Using Ontology", *International Journal of Computer Science & Communication Networks*, Vol. 4, No. 3, pp. 130-136, 2013.

[4] G. Metal, "Personalizing Image Search From The Photo Sharing Website", *International Journal of Research in Computer Applications and Robotics*, Vol. 2, No. 6, pp. 18-23, 2014.

[5] H. Ching, "Integrating ontology technology with folksonomies for personalized social tag recommendation", *Applied Soft Computing*, Vol. 13, No. 8, pp. 3745–3750, 2013.

[6] Z. Zhou, "Social Information Retrieval Based On User Interesting", *Journal of Computational Information Systems*, Vol. 1, pp. 109-189, 2011.

[7] X. Wu, L. Zhang, "Exploring Social Annotations for the Semantic Web", In: *Proc. of International World Wide Web Conference Committee*, pp. 417-426, 2006.

[8] D. Yong, "Enhanced Web Information Retrieval By Topic Tag Mining", *Journal of Convergence Information Technology*, Vol. 6, No. 4, pp. 18-24, 2011.

[9] T. Gruber, "A Translation Approach To Portable Ontologies, Knowledge Acquisition Specification", *Knowledge acquisition*, Vol. 5, No. 2, pp. 199 – 220, 1993.

[10] S. Kalarani, "Integration Of Semantic Web & Knowledge Discovery for Enhanced Information Retrieval", International Journal Of Computer Applications, Vol. 1, No. 1, pp. 99-103, 2010.

[11] V. Devedžić and D. Gaševic, "Web2.0 & semantic web", *Annual of Information Systems*, Springer-Verlag Berlin Heidelberg, pp. 135-165, 2009.

[12] O. Medelyan, E. Frank, and I. Witten, "Human-competitive tagging using automatic key phrase extraction", In: *Proc. of the Internet Conference of Empirical Methods in Natural Language Processing*, pp. 1318-1327, 2009.

[13] P. Gottgtory, N. Kasabov, and S. Macdonell, "An Ontology Engineering Approach for Knowledge Discovery from Data In Evolving Domains", In: *Proc. of the International Conference on Data Mining*, pp. 43-52, 2004.

[14] C. Manning, P. Raghavan, and H. Schütze, "An Introduction to Information Retrieval", *Information Retrieval*, Vol. 13, No. 2, pp. 192-195, 2010.

[15] R. Roul, O. Devanand, and S. Sahay, "Web Document Clustering and Ranking using Tf-Idf based Apriori Approach", *International Journal of Computer Applications*, Vol. 2, pp. 34-38, 2014.

[16] O. Méndez, H. Calvo, and M. Armendáriz, "A Reverse Dictionary Based on Semantic Analysis Using WordNet", *Lecture Notes in Computer Science*, Vol. 8265, pp. 275-285, 2013.

[17] K. Basita, M. Vila, J. Campana, and M. Bautista, "An ontology-based framework for automatic topic detection in multilingual environments", *International Journal of Intelligent Systems*, Vol. 33, No. 18, pp. 1459-1475 , 2018.

[18] A. Rathore and D. Roy, "Ontology based Web Page Topic Identification", *International Journal of Computer Applications*, Vol. 85, No. 6, pp. 35-40, 2014.

[19] M. Rani, A. Dhar, and O. Vyas, "Semi-automatic terminology ontology learning based on topic modelling", *Engineering Applications of Artificial Intelligence*, Vol. 63, pp. 108-125, 2017.

[20] M. Bouadjenek, "Infrastructure and Algorithms for Information Retrieval Based On Social Network Analysis/Mining", *University of Paris-Saclay*, 2016.