

AN EFFICIENT INFORMATION RETRIEVAL ONTOLOGY SYSTEM BASED INDEXING FOR CONTEXT

G.Krishna Raju¹, Padmanabham², A.Govardhan³

¹CS Department, Matrusri PG Studies, Saidabad, Hyderabad, India

²Dean, Bharat Engineering College, Ibrahimpatnam, Telengana, India

³Director, School Of IT, JNTUH, Hyderabad, Telangana, India

Abstract

Many of the research or development projects are constructed and vast type of artifacts are released such as article, patent, report of research, conference papers, journal papers, experimental data and so on. The searching of the particular context through the keywords from the repository is not an easy task because the earliest system the problem of huge recalls with low precision. This paper challenges to construct a search algorithm based on the ontology to retrieve the relevant contexts. Ontology's are great knowledge of retrieving the context. In this paper, we utilize the WordNet ontology to retrieve the relevant contexts from the document repository. It is very difficult to retrieve the relevant context in its original format since we use the pre-processing step, which helps to retrieve context. The pre-processing step includes two major steps first one is stop word removal and the second one is stemming process. The outcome of the pre-processing step is indexing consist of important keywords and their corresponding keywords. When the user enter the keyword to the system, the ontology makes the several steps to make the refine keywords. Finally, the refine keywords are matched with index and relevant contexts are retrieved. The experimentation process is carried out with the help of different set of contexts to achieve the results and the performance analysis of the proposed approach is estimated by the evaluation metrics like precision, recall and F-measure.

Keywords— Ontologies; WordNet; contexts; stemming; indexing.

-----***-----

1. INTRODUCTION

Information Retrieval (IR) deals with the retrieval of all contexts, which contain information relevant to any information need expressed by any user's query. The methodological rule given in literature is to begin an evaluation by analyzing what is the objective of the system, process or service to be evaluated [1] [2]. It is assessed that to what extent the object of evaluation attains the defined goals. Therefore, it is necessary to identify the goals of the system, and measures of goal attainment and criteria for achieving goals [3] [4].

An Information Retrieval System (IRS) consists of a software program that facilitates a user in finding the information the user needs [5]. IR provides the contexts that satisfy their needs. IRS has to extract the key words from the contexts and assign weights for each keyword. Recently, however, researchers have undertaken the task of understanding the human, or user, role in IR [6] [7]. The basic assumption behind these efforts is that we cannot design effective IR systems without some knowledge of how users interact with them. Therefore, the research that studies users in the process of directly consulting an IR system is called interactive information retrieval (IIR) [8] [9].

Query efficiency must be ensured to find out whether the queries are running fast. Query Effectiveness also affects the

IRS since the retrieved result set must be relevant [10]. Research in IR includes modeling, context classification and categorization, system's architecture, user interfaces, data visualization, filtering, languages, etc. A global perspective holds that all of the factors that influence and interact with a user, such as search intermediary, IR system, and texts, should be considered in IR research [11] [12]. The design variables put forth by Ingwersen show the wide-ranging influence of factors such as social environment, IR system, information objects, intermediary, and user [13].

The main assumption is that context does not change in time. However, this assumption is unlikely. Consider Relevance Feedback (RF) technique, the idea behind RF is that the first retrieval operation can be considered as an "initial query formulation" [14]. Some initially retrieved items are examined for relevance, and then the automatic modification of the query can be performed by the system by using the feedback collected from the user for instance adding keywords, selecting and marking contexts [15]. The modified query can be considered a "refinement" of the initial query. A possible solution is the adoption of techniques, which are transparent to the user that is "implicit". Implicit Relevance Feedback (IRF) techniques [16] [17] can use different contextual features collected during the interaction between the user and the system in order to suggest query expansion terms, retrieve new search results, or

dynamically reorder existing results. One of the difficulties with this kind of techniques is the need of combining different sources of evidence, i.e. different contextual features [19]. The complexity of these approaches is one of the reasons for investigating the problem in a principled way that is for the adoption of a model-based development [19]. One of the benefits of this approach is that all the assumptions are made explicit: this is crucial in modeling context in order to understand which elements of the context are actually considered, and in which way the relationship between such elements is modeled [20].

2. RELATED RESEARCHES: A REVIEW

Despite a plenty of works available in the literature, a handful of significant research works are reviewed here. Xuehua Shen et al. [21] have proposed a method for retrieval models and systems that the retrieval decision was made based solely on the query and document collection; information about the actual user and search context. In this proposed method, they studied how to exploit implicit feedback information, including previous queries and click through information, to improve retrieval accuracy in an interactive information retrieval setting. They proposed context-sensitive retrieval algorithms based on statistical language models to combine the preceding queries and clicked context summaries with the current query for better ranking of documents. They used the TREC AP data to create a test collection with search context information, and quantitatively evaluate our models using this test set.

Emanuele Di Buccio et al. [22] have proposed a technique for an information retrieval (IR) system documents according to their predicted relevance to a formulated query. In this proposed method, for each user it is assumed one information need for each query, one location where the user is, and no temporal dimension. Exploiting the context in a way that does not require a high user effort may be effective in IR as suggested. The high number of factors to be considered by these techniques suggests the adoption of a theoretical framework, which naturally incorporates multiple sources of evidence. Moreover, the information provided by the context might be a useful source of evidence in order to personalize the results returned to the user. Indeed, the information need arises and evolves in the present and past context of the user. Since the context changes in time, modeling the way in which the context evolves contributes to achieve personalization.

Massimo Melucci et al. [23] have proposed a method for Information retrieval for context model by vector space base and its evolution was modeled by linear transformations from one base to another. Each document or query can be associated to a distinct base, which corresponds to one context. They proposed to discover contexts from document, query or groups or them. Linear algebra could do thus by employed in a mathematical framework to process context, its evolution and application.

Massimo Melucci et al. [24] have proposed Information retrieval (IR) model based on vector spaces have been investigated for a long time. Nevertheless, they have recently attracted much research interest. In parallel, context has been rediscovered as a crucial issue in information retrieval. This article presents a principled approach to modeling context and its role in ranking information objects using vector spaces. First, the article outlines how a basis of a vector space naturally represents context, both its properties and factors. Second, a ranking function computes the probability of context in the objects represented in a vector space, namely, the probability that a contextual factor has affected the preparation of an object.

David Robins et al. [25] have introduced interactive information retrieval systems. Interactive information retrieval may be contrasted with the "system entered" view of information retrieval in which changes to information retrieval system variables are manipulated in isolation from users in laboratory situations. In this proposed method, they elucidates current models of interactive information retrieval, namely, the episodic model, the stratified model, the interactive feedback and search process model, and the global model of poly representation.

3. PROPOSED METHODOLOGY

Here we proposed a new IR method which is used for recovering traceability links between code and documentation. To access the large database, initially the database will be partitioned by using Jensen-Shannon (JS) method. The JS will be constructed by partitioning the database into smaller sizes. WordNet is an online lexical database of English, developed under the guidance of Miller at Princeton University. Here, a set of cognitive synonyms called synsets, each representing a different concept, are formed by grouping the nouns, verbs, adjectives and adverbs. Synsets are created by using conceptual semantic and lexical relations. WordNet can also be seen as ontology for natural language terms. It has more than 100000 words, organized into taxonomic hierarchies. Nouns, verbs, adjectives and adverbs are grouped into synonym sets (synsets). The synsets are also grouped into senses i.e., diverse meanings of the same word or concept. Same as the Open Directory, the synset ids are altered when new versions of the ontology are published, however a backward compatibility utility program is used to map synsets between the versions.

4. PROPOSED APPROACH OF DESIGN AND IMPLEMENTATION OF AN ONTOLOGY-BASED CONTEXT RETRIEVAL

The aim of this proposed research is to design and develop an approach for the context retrieval by combining keyword in ontology platform. Initially, a user submits the keywords into the system, the ontology operates with the keywords, and a list of contexts is retrieved from the document repository. Initially the system find out the possible synsets (set of synonyms) for

the each keyword as the user entered. Subsequently the system makes the possible combinations of all keywords from the synsets. The neighborhoods are a set of words that are relevant to the combinations. From the collection of neighborhoods, the system count the frequency of the each keyword, if the keyword is supported by the minimum support then the words go to the representation table else the correspond word get neglected. If no words are present after given the minimum _support then the user, get the chance to provide the relevant keyword by the keyword-refining schema. The distance measure helps to find out the refined keywords from the representation. Finally, the keywords are matched with the indexing and relevant context are retrieved.

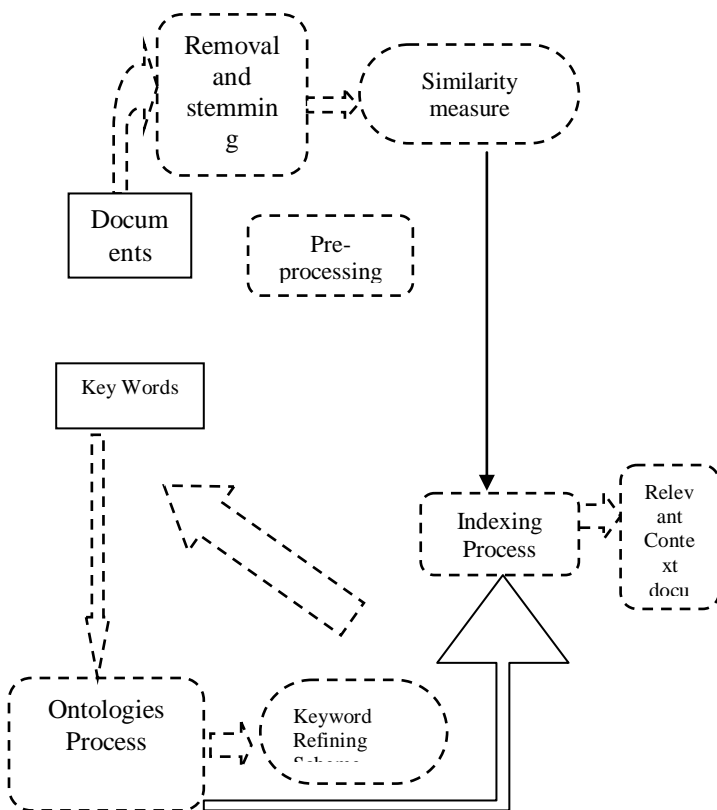


Fig 1: The architectural diagram of the proposed approach

4.1 Preprocessing

In the proposed approach, there are some complexities to deal with the context in its original format since we have to do some pre-processing techniques to make the context repository prepare for our proposed method to retrieve the relevant contexts based on keywords given by the user. The main objective of this pre-processing is to obtain the important keyword from the all contexts present in the database repository. Finding of important keyword from the document repository is not an easy task because each of the context contains a vast amount of common words and branch words. In order to remove those kinds of words from the context the following methods are going to use in the pre-processing phase.

The pre-processing step mainly consists of three steps first one is stop word removing, the second one is stemming algorithm, third one is similarity measure.

4.1.1 Deletion of Stop Words

It is difficult to select keywords in contexts, which have a bulk number of words. Picking the keywords among the huge number of words in a context can be achieved through the stop word removing. The general words (such as was, is, the) are removed through stop word removing process in order to extract the keywords from a context. Because of this procedure, only important words are left as a residue. The major reason of eliminating stop words is to conserve the system resources by deleting those words that have little value for mining procedure. The common words that are noticed as stop words consists of function word and a few more (i.e. articles, conjunctions, interjections, prepositions, pronouns). Stop words like "it", "a", "can", "an", "and", "by", "for", "from", "of", "the", "to", "with" are the common stop words.

4.1.2 Stemming Process

The stemming algorithm has filtered token, this token has the branch words with the root word and this will help to find out the documents, which contain the branch words of the root words. For instance, if a query includes the word walk, the user may desire documents that contain the word walks, walking or walked. This process helps to reduce the need of memory space while indexing process and it helps to make better finding out the relevant documents.

4.2 Indexing

The document retrieval system prepares for retrieval by indexing the documents and formulating the keywords, resulting in document representations and keyword representations respectively. Automatic indexing begins with the important keywords, such as extracting all the words from a text, followed by refinements in accordance with the conceptual schema. After finishing the pre-processing process, the documents contain only the keywords. The system calculates the frequency of keyword in all of the documents. This process gives the count of each keyword. To select the important keywords, the user sets the threshold value. The keyword is preferred as an important keyword when the following condition 1 is satisfied. If the count value of the keyword is greater than the threshold value then that keyword considered as important keyword.

$$imp(K_i) = Cnt(K_i) > Th \quad (1)$$

Here, the indexing is done with the aid of the inverted indexing method so that the matching process can be done easily. The index $I[imp(K_i), D_i]$ consists of important keywords $imp(K_i)$ and their corresponded documents D_i .

4.2.1 Retrieval of Relevant Documents by Using the Proposed Similarity Measure

The user gives the keyword to ontology to retrieve the relevant document, in order to achieve that goal; our proposed similarity measure uses the following approaches. Synset S_i : which helps to find out the relevant synonyms of the keywords as the user entered from the ontology.

A). **Combination**: the combination of keywords helps to find out the relevant neighborhood from the ontology.

B). **Neighborhood**: which helps to find out the nearest words of the combinations, which made in the previous section.

C). **Representation**: which helps to find out the relevant keywords from the collection of nearest words of the combination. If the frequency of the nearest word is greater than the min-support then that word considered as relevant keyword. If there is no words having the frequency greater than min-support then the user has the chance to give the relevant keyword through a keyword-refining schema. Each of the keywords from the representation is matched with the neighborhoods to find out the distance between them. The combination of the keywords selects as a refined keyword one who having the maximum distance among them. Finally, the keywords are matched with the index and relevant documents are retrieved from it.

4.3.1 Synsets

Synset is a set of synonyms of the keyword that are given by the user with the aim of retrieving the relevant document from the repository. User given the set of keywords $K = \{K_i\}$ where $1 \leq i \leq n$ are the keywords to the ontology, the synset of each keyword $S_i = \{K_i(K_{ij}, K_{ik}) \in O\}$ where j and k represents the synset of the keyword K_i , the synset of the keyword K_i is get from the ontology O .

4.3.2 Combination of all Keywords from the Synset

Each of synset have set of keywords each of them are combined with the other keywords for finding the efficient neighborhood for the keyword. The combination of the keyword helps to find out the accurate document through the neighborhood. Every keywords of the each synset combines with other keywords $C_{ij}^m = \{K_{ij}, K_{ik}\}$ where i, j, k represents the keywords in the synsets and m represents the identity of combination.

4.3.3 Neighborhood

The neighborhoods are nearest words of the keyword present in the above of combinations. This will help to improve the quality of the keywords. The neighborhoods are represented for each

combination $C_m = \{(N_p^m, N_q^m) \in O\}$ N_p and N_q are contains the set of nearest words $N_p^m = \{K_{pi}^m\}$, $N_q^m = \{K_{qi}^m\} \in O$ of the keyword K_{ij} and K_{ik} respectively. Likewise, all of the above combinations have the neighborhoods from the ontology.

4.3.4 Representation

Representation is used to find out the important keywords from the neighborhoods of all combinations. Each keyword $N_p^m = \{K_{pi}^m\}$, $N_q^m = \{K_{qi}^m\} \in O$ of the neighborhood of N_p and N_q has the count value. The keywords are removed when their count value less than the min-support that is given by the user. The

representation $R_p = \{R_{pi} \mid \text{Cnt}(K_{pi}^m) > \text{min_sup}\}$,
 $R_q = \{R_{qi} \mid \text{Cnt}(K_{qi}^m) > \text{min_sup}\}$. Finally

R_p and R_q have the set of keywords $R_p = \{K_{pi}\}$, $R_q = \{K_{qi}\}$. If the representation R_p and R_q has no keywords after given the value of min_sup the given by the user, subsequently the user has the chance to provide the relevant keywords through the keyword refining schema.

Pseudo Code

INPUT: keywords, K_i ;

OUTPUT: Relevant Documents R_D

ASSUMPTIONS:

$\text{Cnt}(K_i) \rightarrow \text{Count of Keyword}(K_i)$

$Th \rightarrow \text{Threshold}$

$S_i \rightarrow \text{Synsets of keyword}(K_i)$

$Th \rightarrow \text{Ontology}$

$C_m \rightarrow \text{Combination of Synset and Keywords}$

$(N_p^m, N_q^m) \rightarrow \text{Neighborhoods of Combination } C_m$

$\{K_{pi}^m\}, \{K_{qi}^m\} \rightarrow \text{Set of keywords belongs to Neighborhoods}$

(N_p^m, N_q^m)

$R_p \rightarrow \text{Representation}$

$I[\text{imp}(K_i), D_i] \rightarrow \text{Indexed documents (imp}(K_i) \text{ refer to important keyword and } D_i \text{ refers to corresponding documents)}$

$D_m \rightarrow \text{Distance measure}$

$R_D \rightarrow \text{Relevant Document}$

Pseudo code:*Begin**Step1: Pre-processing of documents by stop word removal and stemming process**Count the Keywords in a set of document $Cnt(K_i)$* *if ($Cnt(K_i) > Th$)**(K_i) denoted as $imp(K_i)$* *else* *K_i removed from documents**Step2: Construct index $I[imp(K_i), D_i]$* *Step3: Get keyword $K = \{K_i\}$* *Step4: Obtain $S_i = \{K_i(K_{ij}, K_{ik}) \in O\}$* *Step 5: Obtain $C_{ij}^m = \{K_{ij}, K_{ik}\}$* *Step6:**Get Neighborhoods belongs to*

$$C_m \subseteq \{(N_p^m, N_q^m) \in O\}$$

Compute representation R_p and R_q

$$R_p = \{R_{pi} | Cnt(K_{pi}^m) > \min_sup\}$$

$$R_q = \{R_{qi} | Cnt(K_{qi}^m) > \min_sup\}$$

Step 7: Compute Distance $D_m = N_p^m \cap R_p + N_q^m \cap R_q$ *Select C_{ij}^m Which has max Value of D_m* *Step 8: Get the Relevant Key Words From $C_{ij}^m = \{K_{ij}, K_{ik}\}$* *$K_{ij}, K_{ik}$ are match with $I[imp(K_i), D_i]$* *Step 9: Get Relevant Document RD_i from $I[imp(K_i), D_i]$* *End***4.3.5 Finding of the Refined Keywords from the Representation**

The keywords from $N_p^m = \{K_{pi}^m\}$, $N_q^m = \{K_{qi}^m\} \in O$ are comparable with the representative $R_p = \{K_{pi}\}$, $R_q = \{K_{qi}\}$ to find out the distance between them. The distance calculation is done by the following equation 2.

$$D_m = N_p^m \cap R_p + N_q^m \cap R_q$$

This distance calculation is done for all the combinations C_m and we get the set of distance measure from that we choose

the distance value D_m which having the maximum value. The corresponding combination is extracted with the help of N_p^m and N_q^m since the neighborhoods are a subset of the combination. The combination has the set of keywords $C_{ij}^m = \{K_{ij}, K_{ik}\}$ that are considered as refined keyword.

4.3.6 Finding of the Relevant Document

The refined keywords K_{ij}, K_{ik} are matched to the index $I[imp(K_i), D_i]$ if the keywords are same then the corresponding relevant documents R_D are retrieved.

5. RESULTS AND DISCUSSION

The results obtained from the experimentation of the proposed cross ontology-based similarity measure for bio-document retrieval system is presented in this section. We have implemented our proposed bio-document retrieval system using Java (JDK 1.6). The dataset utilized in our experimental results are bio-medical documents obtained from the PubMed database.

5.1 Evaluation Metrics

An evaluation metric is used to evaluate the effectiveness of document retrieval systems and to justify theoretical and practical developments of these systems. It consists of a set of measures that follow a common underlying evaluation methodology. Some of the metrics that we have chosen for our evaluation purpose are Recall, Precision and the F-measure.

Precision,

$$P = \frac{\{\text{relevant document}\} \cap \{\text{releventdocument}\}}{\text{retrieveddocuments}}$$

Recall,

$$R = \frac{\{\text{relevant document}\} \cap \{\text{retrieveddocument}\}}{\text{relevent documents}}$$

F- Measure,

$$F = \frac{2PR}{(P+R)}$$

As suggested by the above equations in the field of Document retrieval, Precision is the fraction of retrieved documents that are relevant to the search, Recall is the fraction of the documents that are relevant to the query that are successfully retrieved and the F-measure that combines precision and recall is the harmonic mean of precision and recall.

5.2 Performance Analysis

The performance of the proposed document retrieval system is evaluated based on the input query keywords to the WordNet ontology using the proposed similarity measure. Here, we have utilized four query keywords and the corresponding refined keywords are extracted from the WordNet ontology using the proposed similarity measure. We have analyzed our proposed system with query keywords with the refined keyword. The table 2 lists the obtained values for the evaluation measures with query keywords and the relevant keywords. It reveals that the proposed system works fine in the similarity measure process.

Table 1. Refined keywords for the input query keyword

Query keyword		Max-distance
Query Keyword	Refined Word	
Software processing	development software	7

Table 2. Precision, Recall and F-measure for different keywords

Query keyword		Relevant documents	Retrieved documents	Precision	Recall	F-measure
Software processing	Development software, software documentation	7	10	1	0.8	0.8888
Sequential pattern	Sequential structure	9	10	0.9677	0.8	0.9836
Computer Graphics	host computer, computer graphics, host	10	19	0.8569	0.8	0.8957
Digital Image process	Process	8	10	0.8	0.6	0.6153

6. CONCLUSION

In this paper, we have presented the design and implementation of ontology based document retrieval approach. At first, the set of keywords is extracted from the documents as the outcomes of the pre-processing steps. The indexing process makes the important keyword and their corresponding documents. The refined keywords are extracted by the proposed similarity measure after the user given the input keyword to the system. The refined keywords are matched with the index and corresponding documents are retrieved. Finally, the refined keywords are matched with index and relevant documents are retrieved. The experimentation process is carried out with the help of different set of documents to achieve the results and the performance analysis of the proposed approach is estimated by the evaluation metrics like precision, recall and F-measure.

REFERENCES

[1]. Carla Teixeira Lopes "Context Features and their use in Information Retrieval", the preceding of 3rd Symposium on Future Directions in Information Access (FDIA), pp. 36-42,

Sequential pattern	Sequential structure	14
Computer Graphics	host computer Graphics	16
Digital Image process	Process	18

5.3 Performance Analysis using Evaluation Metrics

The performance of the proposed document retrieval system is evaluated based on the input query keywords to the WordNet ontology using the Precision, recall and F-measure. Here, we have utilized four query keywords and the corresponding documents are obtained from the document repository. We have analyzed our proposed system with different keywords with the relevant and retrieved documents. The table 2 lists the obtained values for the evaluation measures with different keywords and the relevant documents as 20. It reveals that the proposed system works fine in the document retrieving process.

- [2]. Abdelkrim Bouramoul, Mohamed-Khireddine Krolladi, and Bich-Lien Doan "Using Context to Improve the Evaluation of Information Retrieval Systems", International Journal of Database Management Systems, Vol.3, No.2, pp. 22-39, 2011.
- [3]. Ali Bahrami, Jun Yuan, Paul R. Smart and Nigel R. Shadbolt, "Context Aware Information Retrieval for Enhanced Situation Awareness", IEEE Military Communications Conference, Orlando, FL, USA, pp. 1-6, 2007.
- [4]. Peter D. Turney and Patrick Pantel "From Frequency to Meaning: Vector Space Models of Semantics", Journal of Artificial Intelligence Research, Vol. 37, pp. 141-188, 2010.
- [5]. Castells P., Fernandez M. and Vallet D. "An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval", IEEE Transactions on Knowledge and Data Engineering, Vol. 19, No. 2, pp. 261 – 272, 2007.
- [6]. Michael W. Berry, Zlatko Drmac, and Elizabeth R. Jessup, "Matrices, Vector Spaces, and Information Retrieval", Society for Industrial and Applied Mathematics, Vol. 41, No. 2, pp. 335–362, 1999.
- [7]. Gerard Salton and Christopher Buckley "Term-weighting approaches in automatic text retrieval", Information Processing & Management, Vol. 24, No. 5, pp. 513–523, 1988.

- [8]. Vijay V. Raghavan and S. K. M. Wong "A Critical Analysis of Vector Space Model for Information Retrieval", *Journal of the American Society for Information Science*, Vol. 37, No. 5, pp. 279-287, 1986.
- [9]. Faloutsos, Christos, Oard and Douglas W. "A Survey of Information Retrieval and Filtering Methods", *Technical Reports of the Computer Science Department*, 1998.
- [10]. Robert M. Losee "Learning syntactic rules and tags with genetic algorithms for information retrieval and filtering: An empirical basis for grammatical rules", *Information Processing & Management*, Vol. 32, No. 2, pp. 185-197, 1996.
- [11]. Lee D.L., Huei Chuang and Seamons K. "Document ranking and the vector-space model", *IEEE Software*, Vol. 14, No. 2, pp. 67-75, 1997.
- [12]. S. K. M Wong and Y. Y Yao "A probabilistic inference model for information retrieval", *Information Systems*, Vol. 16, No. 3, pp. 301-321, 1999.
- [13]. Marko Balabanovic and Yoav Shoham, "Learning Information Retrieval Agents: Experiments with Automated Web Browning" *AAAI Technical Report SS*, pp. 13-18, 2008.
- [14]. Osinski, S.; Weiss, D., "A concept-driven algorithm for clustering search results," *IEEE Intelligent Systems*, Vol. 20, No. 3, pp. 48-54, 2005.
- [15]. Hayes, J. H.; Dekhtyar, A. and Osborne J., "Improving requirements tracing via information retrieval," *Proceedings 11th IEEE International Requirements Engineering Conference*, pp.138-147, 2003.
- [16]. Holger Billhardt, Daniel Borrajo and Victor Maojo "A context vector model for information retrieval", *Journal of the American Society for Information Science and Technology*, Vol. 53, No. 3, pp. 236-249, 2002
- [17]. Bellegarda J. R. "Latent semantic mapping [information retrieval]", *IEEE Signal Processing Magazine*, Vol. 22, No. 5, pp. 70-80, 2005.
- [18]. Oliveto R., Gethers M., Poshyvanyk D. and De Lucia A., "On the Equivalence of Information Retrieval Methods for Automated Traceability Link Recovery," *IEEE 18th International Conference on Program Comprehension (ICPC)*, pp. 68-71, 2010.
- [19]. Jean Véronis "HyperLex: lexical cartography for information retrieval", *Computer Speech & Language*, Vol. 18, No. 3, pp. 223-252, 2004.
- [20]. Praveen Pathak; Gordon, M.; Fan, W., "Effective information retrieval using genetic algorithms based matching functions adaptation," *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, Vol. 1, pp. 1-8, 2000.
- [21]. Xuehua Shen, Bin Tan and Cheng Xiang Zhai "Context-Sensitive Information Retrieval Using Implicit Feedback", In *proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 43-50, ACM New York, NY, USA, 2005.
- [22]. Emanuele Di Buccio "Modeling the Evolution of Context in Information Retrieval", *the 2nd BCS-IRSG Symposium on Future Directions in Information Access*, pp. 6-12, 2008.
- [23]. Massimo Melucci "Context Modeling and Discovery using Vector Space Bases" *Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM New York, NY, USA, pp. 808-815, 2005
- [24]. Massimo Melucci "A Basis for Information Retrieval in Context", *ACM Transactions on Information Systems*, Vol. 26, No. 3, 2008.
- [25]. David Robins "Interactive Information Retrieval: Context and Basic Notions", *Informing Science Journal*, Vol. 3, pp. 57-62, 2000.