

UML MODELING AND SYSTEM ARCHITECTURE FOR AGENT BASED INFORMATION RETRIEVAL

D. Muhammad Noorul Mubarak, Philomina Simon

Department of Computer Science, University of Kerala
Kariavattom, Thiruvananthapuram, Kerala, India

ABSTRACT

In this current technological era, there is an enormous increase in the information available on web and also in the online databases. This information abundance increases the complexity of finding relevant information. To solve such challenges, there is a need for improved and intelligent systems for efficient search and retrieval. Intelligent Agents can be used for better search and information retrieval in a document collection. The information required by a user is scattered in a large number of databases. In this paper, the object oriented modeling for agent based information retrieval system is presented. The paper also discusses the framework of agent architecture for obtaining the best combination terms that serve as an input query to the information retrieval system. The communication and cooperation among the agents are also explained. Each agent has a task to perform in information retrieval.

KEYWORDS

Intelligent Agents, Crawling, Agent based information retrieval, Object Oriented modeling, Unified Modeling Language, Ontology, Agent Architecture

1. INTRODUCTION

The web contains abundant data and it is getting updated on an hourly basis. Usually, the retrieved results are not relevant to the input query. There is a drastic need for improved search strategies in an information system. Current search engines are primarily passive tools. The retrieval of the information has become a tedious task now days because of the enormous amount of structured and unstructured data. Web users find it difficult to retrieve a relevant result based on the input query. The relevant information can be obtained in the form of textual documents, video files and audio files and so on. The use of intelligent agents [4] in information retrieval for efficient search and retrieval techniques. Lot of research works are going on in the area of information retrieval and intelligent agents to obtain more relevant information.

In our method, the relevant information is obtained in the form of textual documents. Documents are represented by document descriptors. Document descriptors are defined as keywords or terms in a document collection. The system architecture is proposed to derive the combination terms ie, best query terms. Combination terms are obtained from the list of keywords in the document collection. Combination terms are keywords which help to frame the query terms which doesn't have a related meaning. For example, "Genetic Algorithms for Information Retrieval" [8] here, genetic algorithm and information retrieval are not co-occurrence terms and hence they do not belong to the same class. Co-occurrence terms are keywords which have a related meaning. Hence use of co-occurrence data for document retrieval is not always considered as an efficient method. This method gives limited performance when the keywords have unrelated keywords. To overcome the limitations, a method for finding combination of terms is proposed. This concept of

combination terms work better in document retrieval since the keywords does not have any related meaning.

The proposal describes an approach in which an intelligent agent can be incorporated with information retrieval to retrieve the relevant information.

A. Intelligent Agents

In agent based architectures, agents act as mediators between user specific task and a particular application [6]. They are often decomposed into functional modules that refer to perception, modeling, planning, coordination, and task or plan execution. The strong point is that the agents provide the ability to execute and coordinate independently and autonomously sub-tasks, which lead to completing the major user task. In this context, computers have to assist users in planning, coordinating and delegating a specific task. Software agents play a crucial role in the human-computer interaction and performing the coordination of the processes of the retrieval system.

Based on the functionality, Agents have different goals to perform based on the context and the application. Conceptual structures enhance the functional operationalism of the agent by applying it for knowledge representation and its manipulation mechanisms, navigation, and problem definition. Ontology plays a key role in describing and defining a particular context. Agent-based intelligent systems can provide an enhanced learning and understanding for a specific functionality of a system.

Intelligent agents have different features such as social ability, proactiveness and task oriented and task delegation behavior. The agent itself collects the request from the user and the task delegation is performed. The user who searches information from the document collections is an agent. The user agent may teach the intelligent agent by example or by employing a set of criteria for the intelligent agent to follow. Intelligent agent sense and act autonomously in its environment. To some extent, intelligent agents are designed to be adaptive to their environments and to the changes in their environments. Three major approaches for building agents for the document collections are discussed. Firstly, integrate agents into existing search techniques. Secondly, intelligent agent follows rule based filtering approaches. With this approach, an agent is given information about the application. A knowledge engineer generates the rules and knowledge for the agent. Thirdly, In the training approach, the agent collects the information around it and actions of its user (Jensen, 2002).

This proposed work describe the architecture of an intelligent agent that searches the document collection and generate a combination terms which serve as an input query to the information retrieval system.

The paper is organized as follows. Section 1 explains the introduction to information retrieval. Section 2 discusses the different strategies and methodologies for performing information retrieval. Section 3 presents the proposed architecture for Agent based Information Retrieval. Section 4 deals with the cooperation and communication among different agents. Section 5 describes the UML modeling of the agent based system architecture. Section 6 presents the results and discussion. Section 6 summarizes and concludes the paper.

2. LITERATURE REVIEW

In this section, a study on different agent based information retrieval is performed, also the survey of different strategies and methodologies for performing information retrieval is also presented in Table 1.

SAIRE [2]-A Scalable Agent-Based Information Retrieval Engine.

It is a multi agent based search system which incorporates agents, language processing and heuristic search techniques to support Earth and Space Science data over the web. Agents in SAIRE form seven modules, in which an agent manager controls several local agents. Agent managers communicate and send messages through master agent which act as a central coordinating agent. If the central agent fails, then it may have huge impact on other agents and it may lead to system failure. A natural language parser extracts a frame containing actions and concepts from the natural language input. Two dictionaries are used to specify a user's domain of interest. SAIRE performs several functions such as communication activities, directory services, knowledge archiving and retrieval services, data storage access services and information processing services. SAIRE is a Goal-oriented, autonomous intelligent software that is developed to achieve intelligent information retrieval.

SARI [1] The SARI (Software Agents for Retrieval of Information) system act as a mediator between human users or computers at one end, and heterogeneous information sources with various search engines at the other. In SARI, different agents work corporately and coordinately to perform a specified task. Application Agents represent the users or stake holders to the SARI system. They send agent messages for performing information retrieval to Control Agents. Search Agents gather queries from Control Agents and send the results back to the Control Agents. Control Agents act as mediators in the SARI system. Control Agent receives agent messages containing query requests from Application Agents. Search Agents forwards the requests, sends messages containing the retrieval request and receives search results from the Search Agents, incorporates them into information retrieval results and retrieved results are sent to Application Agents. Ontology Agent contains metadata that describe the structure of the information present in the information sources.

MARS [3] (Multi-Agent Retrieval System) is discussed to test the integration of multiple retrieval methods in a single system for more user flexibility and complementary retrieval. The retrieval method incorporated in MARS is an intelligent conceptual retrieval based on the use of a thesaurus, semantic retrieval and conceptual browsing. The system is developed in client-server architecture. In the system architecture, the telecommunication facilities enables distributed agents to exchange messages, standards for message formats and server options and a name server so that agents can locate each other and also the security considerations. Since term co occurrences have disadvantages, the author used the semantic retrieval strategy. Concept retrieval is used in this system compared with keyword retrieval, since it offers more flexibility to restrict the retrieval by replacing a concept in a query. With the rapid growth of information and multiplicity of retrieval techniques, it is neither reasonable to store all information in a single database nor gather all evaluation methods in a single program.

XSAR [5] a dynamic, DTD independent search agent. XML repositories help to generate structured data so that we can process easily. Bajaj et.al presented a novel search tool: XSAR (XML Based Search Agent For Information Retrieval) to access the information repositories. XML build the information in a structured manner much like in a relational database. Now a days XML format is the best way of sending information content across web. Agents access information and XSAR reformulates the input query as an XQL representation, and then launches a spider, a crawler that traverses the database repository and executes the XQL query for each XML page in the database.

Table 1. Survey on different strategies for Information Retrieval

Authors	Type of information retrieval	Methodology	Purpose	Evaluation Metrics	Comments/ Features
R. Dhanapal (2008) [10]	Document Information Retrieval	Edge Index Graph based	Perform web document clustering and contextual indexing	Indexing time	Scalable , Dynamic and Contextual retrieving system
Suriati Akmal et.al (2014) [11]	Product Information Retrieval from Web	Ontology based Formal concept analysis (FCA) used to design ontologies	Generate ontology-based Semantic Similarity measures	Feature based Similarity Measure Composite similarity	Efficient
Shaw et.al (2002) [12]	Integrated Intelligent Information Retrieval for Distributed Heterogeneous Data sources	Agent-based architecture	Design of an Intelligent Information Retrieval and Filtering system	Intelligent system is developed and tested	Modular Flexible and Scalable
Seyfi et.al (2016) [13]	Document Information Retrieval based on Focused Crawler	T-Graph Principles and Scoring function	Finding relevant documents and prioritizing them for effective download	Overall Similarity Measure Priority Score	Accuracy of prediction of an unvisited page Reduce processing overhead
D. Yoo (2012) [14]	Personalized Information Retrieval on the Semantic Web	Hybrid Query Processing Method on Semantic Web	Querying for inferred knowledge base,	Processing time	Quick response time Efficient retrieval of personal information.
Addis et.al (2012) [15]	Information Retrieval with X.MAS	Multi agent systems	Retrieving, Filtering and Reorganizing information according to user interests	Degree of pairwise similarity under the threshold of 25%.	Effective in the design and concrete realization of IR applications
Ropero et.al 2012 [16]	Fuzzy Logic Intelligent Agent for Information Extraction	Fuzzy Logic-based Term Weighting Scheme	Intelligent agent to manage any set of knowledge where information is vague or imprecise	Web page identified by questions Term weighting	Flexible Better Extraction of information
Pai et.al (2013) [17]	Semantic and Ontology based information retrieval	Semantic-based Content Mapping	Extracts semantic features and structure of the documents to form a content	Precision Recall F-Measure	Improved accuracy

			map and then performs the search based on this.		
Khan et.al (2012) [18]	Instructional Video Information Retrieval	E-Learning management systems with Natural Language Processing	Semantic engine helps to derive the meaning of the input words	Question and Answer Module	Good understandability of the user content
Belsis et.al 2014 [19]	Agent architecture benchmark	Agent based	developed implements a distributed knowledge base management system	Round trip time Round Trip delays	Test bed for benchmarking and performance monitoring JADE used
Teng-yang et.al 2012 [20]	Ontology based information retrieval	Vegetable e-commerce ontology	Domain ontology helps to index information and information inference.	Recall Precision	Usage of Semantic Information Annotation model
Attia et.al[21] 2014	Fuzzy Information Retrieval Model	Linguistic based multi-view Fuzzy Ontology Information Retrieval model	Rranking algorithm that ranks the set of relevant documents	Document matching degree	Linguistic based user query system

3. PROPOSED ARCHITECTURE FOR AGENT BASED INFORMATION RETRIEVAL

Even though searching several types of information in the internet, user is not satisfied with the results obtained. Here a different approach is proposed to overcome the limitations of information retrieval. So this concept of combination terms [7][8] is useful for retrieving those documents from the document collection in which the query term does not have any related meaning. The different agents and their features is discussed in the following sections. Crawling agent automatically gathers the keywords and documents which is available in the document collection. The indexing agent collects the information from the crawler agent and stores into database. Search agent searches for the more relevant data in the database and it retrieves more relevant results based on the combination of terms. Indexing agent indexes and stores the information which is collected from the crawler. Query agent processes the query send by the user agent and this agent delegate the search task to search agent. Retrieval agent checks that the pages retrieved are relevant in nature. If the user query and the retrieved results are matches by the retrieval agent, then the results are retrieved. Combination agent: learn from the keywords stored by the index agent and generates a combination of the keywords which does not have any related meaning.

The components involved in the agent based information retrieval are Crawling agent, Search agent, Indexing agent, Query Agent and user agent. A crawler is a type of software agent. Crawling agent is used to crawl through the document collections and extract the keywords

A. Ontology

Ontology represents the formal naming, definition of underlying concepts, properties and relationships among different modules in a system design. Ontologies help to perform communication and perform decision tasks in relation to the information stored in the document collection. Ontologies are also used to represent the syntax, semantics and the messages exchanged among agents. An ontology is a set of classes, entities, properties and relationships, describing a standardized terminology (W3C, 2009). Users apply ontology in all applications sharing domain specific knowledge. Communication is possible only if agents share a common ontology, or if there are associations between logically separated ontologies (FIPA, 2000). In our approach, two ontologies are discussed. First, Document Ontology which represents the syntax, semantics and relationships of document collections is discussed. Second one is domain-specific ontology which deals with information retrieval and logical relationships between the modules. Domain ontologies, which represent the content and the specific structure of data sources connected to the system. Document ontology deals with the keywords extraction from the document crawler and finds the best combination terms from a set of keywords extracted. Combination terms represents the terms which does not have related meaning.

4. COOPERATION AND COORDINATION AMONG AGENTS FOR INFORMATION RETRIEVAL

The proposed architecture for the agent based information retrieval is shown in Fig 1. The crawler agent gathers and collects information from document collections. The intelligent agent automatically collects the information and keywords and sends that to indexing agent. The crawler agent is so flexible to collect any type of textual information as per the user request given by the user agent.

The index agent collects the information from the crawler agent and stores into database. The index agent collects different types of information such as documents, keywords, hyperlinks and URLs. So the database consists of the above information. The user agent processes the user request. This request is passed to the crawler agent. The concept of the task delegation, goal oriented nature and flexibility is well inbuilt in this proposed architecture. The agents interact intelligently and autonomously to fulfill the functionalities delegated to them.

The combination agent generates a combination of different keywords and these combination keywords are stored in the database. The best combination terms are generated by applying the genetic algorithm [8]. When the user agent sends a query, the query agent forwards the query to search agent. Search agent performs the search operation and it is send to retrieval agent. Retrieval agent analyses the search results and gives the result to the user agent.

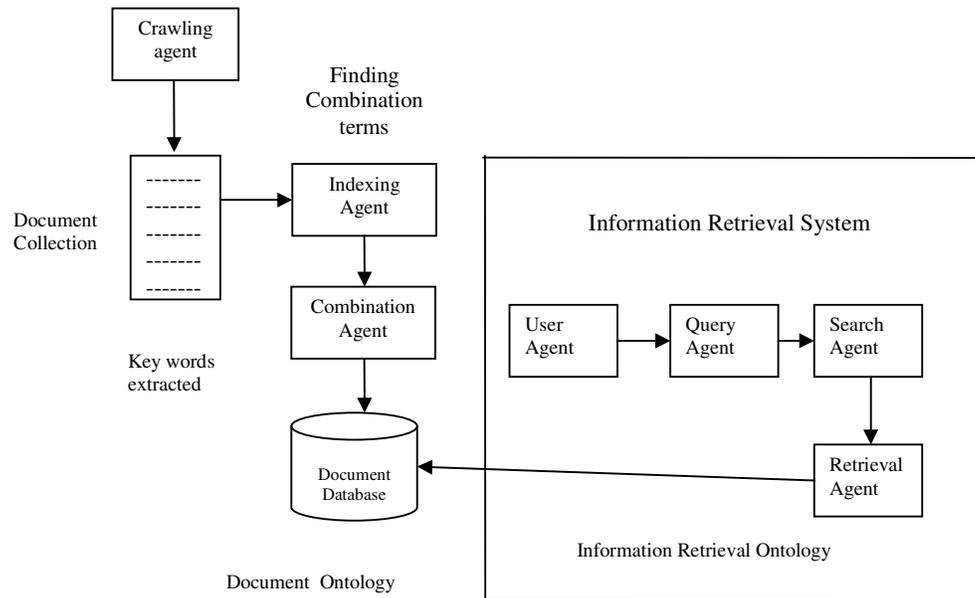


Fig1. Proposed Architecture for retrieving documents using intelligent agents.

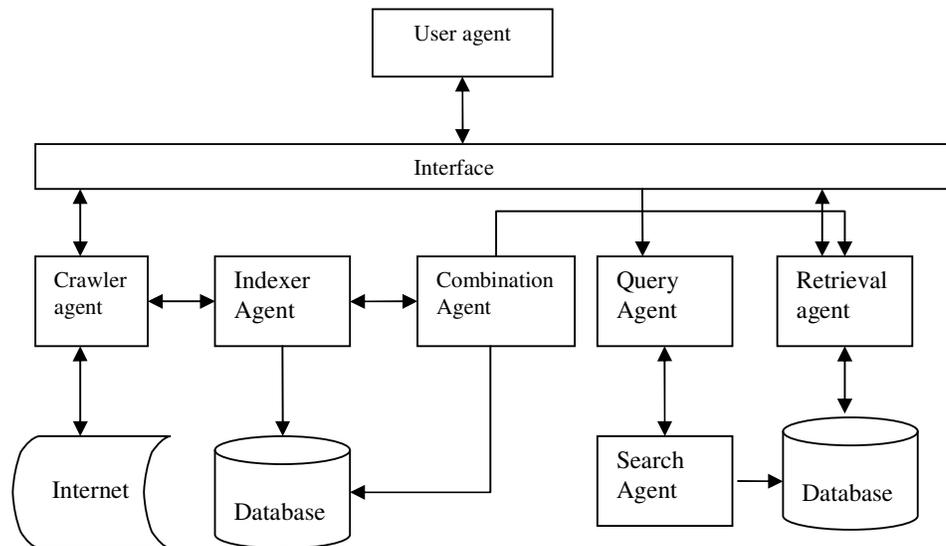


Fig 2: Agent Communication: Cooperation and Coordination

5. UML MODELING FOR AGENT BASED INFORMATION RETRIEVAL

The components involved in the sequence diagram are Crawling agent, Search agent, Indexing agent, Query Agent and user agent. The sequence diagram is represented in Fig 3. The UML representation for the Search agent is shown in Fig 4.

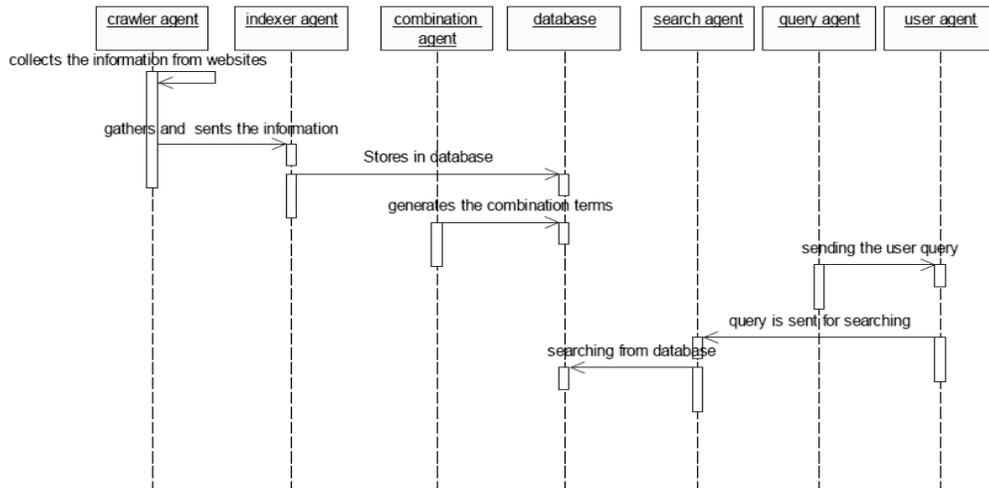


Fig 3: Sequence diagram for Agent based Information Retrieval System

6. RESULTS AND DISCUSSION

The challenges that an information retrieval system face is that they retrieve some irrelevant documents or they do not retrieve all the relevant documents. The performance of the information retrieval system can be calculated by two evaluation measures: Precision and Recall. Precision is defined as the proportion of the retrieved documents that are actually relevant. Recall is defined as proportion of all relevant documents that are actually retrieved. If searchers want to raise precision, then they have to narrow their queries. If searchers want to raise recall, then they broaden their query. It represents many expected docs. The input given is a set of documents. The program for the crawler is tested to extract different number of documents and the keywords are stored. From the keywords, combination agent generates terms of different combinations and it is stored separately. When the user input a search query, the search agent retrieve the relevant set of data that is best suited for the input query. The system architecture for agent based information retrieval is implemented.

7. CONCLUSION

Agent architecture and object oriented modeling for document based information retrieval is proposed. A review on different agent based information retrieval systems is discussed. In the proposed system, combination agent generates a good number of probable search queries and save it in database. This will serve as an input query for the information retrieval system. A system is developed which generated the best combination terms for query processing and information retrieval is performed.

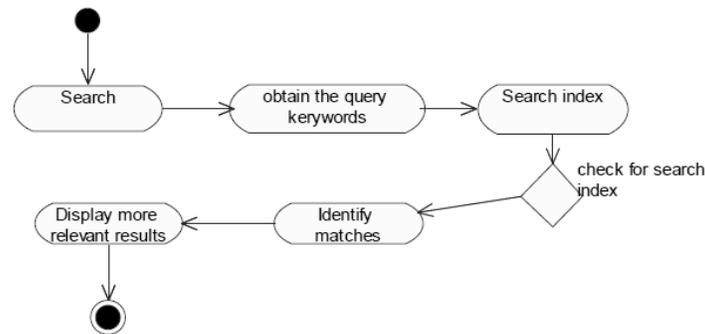


Fig 4: Search Agent: Activity diagram

REFERENCES

- [1] Kuldar Taveter ,”SARI- System for Semantic Information Retrieval”, ERCIM, Advanced databases and Metadata, 1998
- [2] Jide.B.Odubiyi, David J.Kocur et.al, SAIRE- a scalable agent based information retrieval, Proceedings of First International Conference on Autonomous agents,ACM, 292-299,1997
- [3] J.Nie,Anquetil,J.Vaucher,”A Multi-Agent Distributed Retrieval System”, CAIS-ASCI Conference Proceedings ,1993
- [4] Masoud Mohammadian ,”Intelligent Agents for data Mining and Information Retrieval”, 2004
- [5] Bajaj, Akhilesh; Tanabe, Hideaki; and Wang, Chao-hin, "XSAR: Xml Based Search Agent For Information Retrieval" (2002). *AMCIS 2002 Proceedings*. Paper 197
- [6] Lora Aroyo , Piet Kommers,”Intelligent agent for educational and computer-Aided systems”, Journal of Interactive learning Research ,Special issue., 10(3/4), 235-242, 1999.
- [7] Philomina Simon, A Two Stage approach to Document Retrieval using Genetic Algorithm, International Journal of Recent Trends in Engineering, Vol 1, No 1, 526-528, 2009
- [8] S.Siva Sathya , Philomina Simon, A Document Retrieval System with Combination Terms Using Genetic Algorithm, International Journal of Computer and Electrical Engineering, Vol. 2, No. 1, pp 1793-8163, February, 2010
- [9] [http:// www.w3.org/ TR/ rdf-sparql-query/](http://www.w3.org/TR/rdf-sparql-query/) **W3C. (2009)**
- [10] R. Dhanapal,"An intelligent information retrieval agent", Knowledge-Based Systems 21 (2008) 466–70
- [11] Suriati Akmal,Li-Hsing Shih, Rafael Batre, "Ontology-based similarity for product information retrieval", Computers in Industry, Vol 65, Issue 1, Pages 91-107, 2014
- [12] Neal G. Shaw, Ahmad Mian, Surya B. Yadav,"A comprehensive agent-based architecture for intelligent information retrieval in a distributed heterogeneous environment", Decision Support Systems 32 (2002) 401– 415
- [13] Ali Seyfi, Ahmed Patel,"A focused crawler combinatory link and content model based on T-Graph principles", Computer Standards & Interfaces 43 (2016) 1–11
- [14] Donghee Yoo,"Hybrid query processing for personalized information retrieval on the Semantic Web", Knowledge-Based Systems 27 (2012) 211–218
- [15] Andrea Addis, Giuliano Armano, Eloisa Vargiu,"Multiagent systems and information retrieval our experience with X.MAS",Expert Systems with Applications 39 (2012) 2509–2523
- [16] Jorge Roperro, Ariel Gómez, Alejandro Carrasco, Carlos León,"A Fuzzy Logic intelligent agent for Information Extraction: Introducing a new Fuzzy Logic-based term weighting scheme", Expert Systems with Applications 39 (2012) 4567–4581
- [17] Mao-Yuan Pai a, Ming-Yen Chen b, Hui-Chuan Chu c, Yuh-Min Chen, "Development of a semantic-based content mapping mechanism for information retrieval", Expert Systems with Applications 40 (2013) 2447–2461

- [18] Emdad Khan, Adel AlSalem, "Ivia: Interactive Video Intelligent Agent Framework for Instructional Video Information Retrieval", International Educational Technology Conference, Procedia - Social and Behavioral Sciences 64 (2012) 186 – 191
- [19] Petros Belsisa*, Stefanos Gritzalisb, Costas Lambrinouidakisc, Christos Skourlasa, Vassilis Tsoukalasa, Dimitris Vassisa, "An Agent Based Architecture Benchmark", ICININFO, Procedia - Social and Behavioral Sciences 147 (2014) 429 – 435
- [20] TAO Teng-yang and ZHAO Ming, "An Ontology-Based Information Retrieval Model for Vegetables E-Commerce", Journal of Integrative Agriculture 2012, 11(5): 800-807
- [21] Zeinab E. Attiaa, Ahmed M. Gadallaha and Hesham M. Hefnya, "An Enhanced Multi-View Fuzzy Information Retrieval Model based on Linguistics", 2013 International Conference on Applied Computing, Computer Science, and Computer Engineering , IERI Procedia 7 (2014) 90 – 95