

ID 642

# A Review of Agent-Based Frameworks for Information Retrieval

G Gayamini# and NT Jayathilaka

Department of Computer Engineering, General Sir John Kotelawala Defence University, Sri Lanka
#gayamini.g@outlook.com

**Abstract**—With the development of the Internet in the digital age, operative technologies that utilize automated tools for searching and retrieving information in any domain, even those not on the web, are in great demand. However, the enormity of the World Wide Web (WWW) poses a challenge for researchers to retrieve useful and precise information to meet their requirements. An Information Retrieval (IR) system is meant to form a stored knowledge base, with items accessible to the information seeker. A major problem of the traditional IR systems is their inability to provide users with a semantic description of the knowledge needed by them. This problem is addressed by this Intelligent Information Retrieval (IIR), which is capable to give much more relevant and accurate information. The need to discover and observe the real-time mutations in knowledge and information requires new techniques in the web IR process. The results of IR contain an abundance of information that matches with the queries or searches in varying degrees of relevance. The relevance of the results is an important concern and often associates with the volume of the results: the bigger the volume of information, the better the relevance, while a lesser volume of information may have less relevant content. Seeking solutions for this issue makes Web IR an active and interesting domain of research and development. Considering the past two decades, interest among many has arisen in software agent technology and its applications. With Intelligent autonomous agents being most suitable for numerous applications in a semantic web environment, many researchers have proposed different frameworks, which comprise of details such as information collecting agents, storing agents, reasoning agents and querying agents. These structures often take into consideration semantic web and intelligent agents research, and other technologies such as information retrieval and knowledge modeling. This study focuses on a brief survey of Agent-

based IR Systems on semantic web and ontology. The performance of such intelligent systems is calculated by considering the productiveness, quality of the search and the results obtained, time performance, and whether users are satisfied with the search results.

Keywords: intelligent information retrieval, intelligent autonomous agents, world wide web, knowledge modeling

#### I. INTRODUCTION

About two decades ago, humans fulfilled their requirements for information and knowledge through books. As the days went by man's search for more information and knowledge started to increase in scope and regularity. Following this, the world began to see the rapid development of technology and the Internet which grew into an environment of knowledge in which various information was incorporated in a more complex and spontaneous manner. In this digital era, millions of people use this developed Web to retrieve information often. According to an analysis conducted by the International Telecommunication Union in 2015, the Internet users per 100 people was as below.

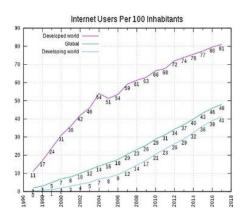


Figure 1. Analysis of Internet Users (*International Telecommunication Union*, 2015)



It is much difficult to provide effective search tools for information access with such large information which is available on the World Wide Web. This leads to developing methods for automatic information retrieval from the appropriate information sources. Traditional information retrieval (IR) methods typically generate database codes for each piece of information on one platform and use popular retrieval models such as the TFXIDF-based vector space model (Choi and Yoo, 1999). Later many approaches were presented to improve the relevance of the information, commonly known as the semantic web.

The drawbacks such as the rapidly changing information, information retrieval speed and limited search range on the Internet, along with many other challenges that came during crawling, indexing of web pages in earlier days were effectively handled by the web search engines(Bargain, 1999; Choi and Yoo, 1999; Prasad Kantamneni and Narayanan, 2001; Rajendran and Balasubramanie, 2009; LUO Junwei, 2010). Users who are not familiar with search engines also use the query interface and Natural language queries to handle problems of information retrieval such as uncertainty, incompleteness, etc. This led to weak navigation and coherence in information retrieval, hence causing flaws in providing users with specific search results depending on their interests. Many systems are being developed to address this challenge in obtaining and filtering specific information. Thus, the distributed indexes or distributed engines can lead to avoiding the bottleneck problem. In 1999(Choi and Yoo, 1999), the mobile agents and an interface were proposed as a solution for this problem where the Mobile Agents (MAs) commonly known as autonomous programs, search the network on behalf of their owners while interacting with other agents for action on the IR process for information.

#### II. DATA AND METHODOLOGY

This section will discuss in detail, the methodology and the approach taken in conducting the review and coming forth with the final paper.

A Systematic type of approach was adopted when writing this review, where the area of research, focus, objectives, and title were decided upon first. The research articles, resources and documents were searched for and selected next, out of which some resources were then used to get a better idea and understanding of the technologies, concepts, and theories. The research papers and articles found were sorted and analysed to select a few successful works on relevant systems to review further. These works were then further studied and reviewed thoroughly, to retrieve the most suitable and appropriate data and information for this research. This information was then utilized to carry out the review to compare the systems, hence resulting in the conclusion as can be seen in this paper. The figure below depicts the workflow of the writing process of this paper.

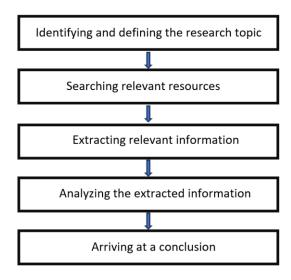


Figure 2. Methodology for the systematic literature review

The disciplines for the study were chosen as Agent-based applications and Intelligent Information Retrieval based on the research topic 'Agent-based Frameworks for Information Retrieval' and the keywords, Intelligent Information Retrieval, Intelligent autonomous agents, World Wide Web, and Knowledge modeling were decided upon relevance. To find the most appropriate and recent projects, research and developments done approximately within the past decade were chosen.

Various resources were referenced to search for the knowledge, information and past research required for this study. Online sources such as Google scholar, and other research archives as



well as resources such as books and repositories were used in information gathering.

The following criteria were considered when selecting the research and articles for this study.

- Works being in English and understandable
  - Entire paper or article being accessible
- Research being within the period considered

Concerning these criteria works such as research papers, case studies, review papers and some conceptual papers that are useful for this research were found. After reviewing the titles and abstracts of these papers, 54 papers were selected as the final works to be referenced for this review paper.

These selected papers were then thoroughly studied, reviewed, and analysed, to understand their suggested systems, solutions, their advantages, and drawbacks as well as the technologies and approaches adopted. 10 – 12 systems such as the information retrieval system by LUO Junwei (2010), and the multi-agent framework by Shudong Zhang, Ye Qin and Naiming Yao (2010) were used for the analysis and comparison.

Factors including the techniques or areas of technology incorporated in the systems, the results of each Information retrieval system and the challenges or drawbacks of each system were analysed and extracted for comparison as in Table 2. The systems were also observed for the types of agents used in each chosen research, where the presence of certain agents such as User, Mediator, wrapper, mobile agents, and more were considered and compared as in Table 1.

The information and knowledge obtained by reviewing, analysing, and comparing these research works were finally utilized to arrive at a few conclusions regarding the traditional as well as the most recent agent-based IR systems and the technologies that are most effective in retrieving relevant information efficiently.

#### III. LITERATURE REVIEW

This section illustrates the researcher's developments for the information retrieval

frameworks adopted by agent technology. The term document is widely used in articles to include any text in a machine-readable format. In the past works, the systems were not able to provide specific search results according to user's interests. Most searching mechanisms are based on keywords or phrases which browsed, so then the content cannot match with the requirement of IR. Moreover, these searching techniques are not providing the information according to relevancy or interests.

In each case, the IR system is designed to extract the best-suited products to meet the needs of the user in response to their demands. The IR system can be used to access drawings, audio and video files, photographs of museum artifacts, patents and more.

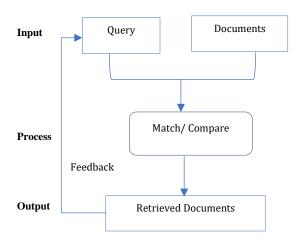


Figure 3. Historic model of an IR system

The above figure 2 illustrates a typical model of the IR system. Researchers have designed a framework (Choi and Yoo, 1999) that is a static multi-agent system for the IR on heterogeneous distributed sources, which fundamentally consists of User, broker, and resource agents, as its components. An electronic shopping agent named as BargainBot (Bargain, 1999) is structured to browse the relevant product or items on the internet.

Research (Travis Bauer and David B. Leake, 2001) has suggested a multi-Agent framework for Handling Complex Information Environments. The name of the system is Calvin. This system observes users as they access documents and proceeds to find related and relevant documents to then provide a unified



interface to the information environment. Calvin is a system built on top of a new Java middle-ware system named 'Geneva,' which is a genebuilttiagent system framework that supplies the resources required to support personal IR. This middle-ware system Java serves with functionalities such basic agent communication, authentication, and encryption functionality, and utilizes open Extensible Markup Language (XML)specifications for agent communication.

Another work (Xiao, 2007) proposed the framework for intelligent retrieval model which deploys a Resource Description Framework (RDF) Model to recognize the web resources to performing the tasks of web modelling and content representation. For RDF data storage and inquiries, the design is adapted to the corresponding database program according to the RDF data model, and XML data can also be used according to the characteristics of the RDF. The comprehensive semantics provided by the Web Ontology Language (OWL) will enable knowledge modelling in specific domains. It provides a knowledge base (KB) for semantic-based reasoning.

A blueprint for Intelligent and Dynamic Business Information Retrieval adopted by the agent concept was presented by Hua Hu, Bin Xu in 2007. The purpose of the A2DT project is to build the framework for the retrieval and analysis of information of dynamic businesses using agentbased technology. This framework which constitutes of agents performing detection, collection, generation, and propagation of information, and policymaking in the framework was established in the A2D project. The authors stated this blueprint is much faster than traditional retrieval. Further, the dynamic retrievals of dynamic results resulting in better performance were highlighted as advantageous. It was mentioned how dynamic information retrievals make the system an intelligent and dynamic process model. However, they faced issues with dynamic retrieval use more storage, and contain more low relevance rank data.

An investigative study on a new information retrieval system incorporating Semantic Web with Multi-agents (LUO Junwei, 2010), analyses the system's ability to process, recognise, extract, extend and match content semantics to attain the

following objectives: (1) integrate a Resource Description Framework (RDF) to analyse and understand the semantic features of the queries, to derive a newer algorithm in order to retrieve the semantics in the context and build up the database consisting semantics; (2) to build a matching algorithm with the use of the semantics retrieved from the content, to be feedback as appropriate, precise and accurate information to satisfy the users' requirements; (3) to bring forward a new Multi-agent based Information Retrieval method, where the Agents within the new model can utilize the users' interests, collect useful information depending on users' behaviours, and extract semantics from the internet to feed it back to distribute this information between many different users, so that the system can retrieve much more accurate that align well with information requirements of the users and can serve the users in completing complex functions and tasks.

In 2010, researchers developed (Shudong Zhang, Ye Qin and Naiming Yao, 2010) a framework that uses agent technology to extract multi-slot web information. The intention of this multi-slot extraction mechanism while addressing information extraction rule learning and repair is to boost the adaptability of this system. Based on the structure, agents are classified and designed for the user, intermediary, wrapper, data store and page pre-processing. where each of these agents incorporates their own Knowledge Bases (KB), the KB classified considering user, PageRank, extraction, tractrack, URL, respectively. Throughout the process of this framework, the user agent retrieves and optimizes the user queries and requests to then forward them to a mediator. The agent for the mediator chooses the URL seeds relative to the requests and hands over the queries and URL seeds to a wrapper. The Web resources of this framework are categorized by page preprocessing agent. The wrapper agent obtains the results extracted from the pages through methods such as data extraction, data tidying, etc. These results are finally returned to the mediator agent where they are integrated and submitted to the user agent for presentation. Furthermore, the author proceeds to point out the better adaptability and expansibility of this model.



An investigation was done with more than eight autonomous agents by Shakti Kundu in 2011. That design was built to be used in agents-based eCommerce applications for web data mining. They implemented an agent that functions to identify vital trend mutation as well as new and emerging information. They appear to have been considerate when Collect user registration information so that user privacy is not infringed through the web data mining process, hence making web data mining with an intelligent system their proposition for future development.

An Author (EFFANGA et al., 2011) constructed a Java Agent Development Environment (JADE) mobile agent system, which proposed a webbased Graphical User Interface (GUI)framework for the JADE mobile agents to show the states of comprising of retrieved platform, information that is stored in a database. This mobile-agent software incorporates a mobile agent platform (executable environment), a mobile agent program (code), and a program based on GUI. This system can perform operations such as Creating, Cloning, Dispatching or Migration, Retraction, Activation, Deactivation and Disposal. At the beginning level, the system was executed with some technical issues such as problems in deploying mobile agents which concerned security, privacy, relireliability, integrity.

A structure (Menacer and Guenaoui, 2016) proposed an approach designed using integrated mobile agent technology for IR on www. Mobile agents can be used securely through a market-based framework to reveal user queries and execute distributed index.

### M. Types of Agents

This section illustrates the types of agents that were developed to perform Agent-based Information Retrieval tasks from the studies mentioned in the literature review. There are some agents commonly implemented in several studies. At the same time, some researchers described their framework by using only mobile agents. Have a deep look into the multi-agent topic all are agents designed to perform some specific tasks allotted to them.

### 1). User Agent

This is the special agent designed to interact with the end-users. Commonly the job of a User-Agent (LUO Yingwei, 2003; Hu and Xu, 2007; Liu *et al.*, 2007; Xiao, Xiao and Zhang, 2007; Shudong Zhang, Ye Qin and Naiming Yao, 2010) comprises environmental concept, memory base, knowledge base, learning of machine and inference of machine, which means it provides users an interface to provide them with definite query requests and to extract information, to then submit these queries. But some of the frameworks have separate agents to carry out procedures such as information and query.

### 2). Search Agent

These agents are assigned to find and locate relevant information for negotiation by the users' requests. In other studies, the author indicates them as information gathering or collecting agents.

### 3). Storing Agent

The metadata from agents collecting intelligence and information is restored and stored properly, by these agents.

### 4). Retrieving Agent

These agents generate matching algorithms to enable a fast process of matching and searching for suitable content. They consist of constituents such as Knowledge Base, Model Base, and Semantic Matching.

#### 5). Mobile Agents

Mobile agent-based technologies have been incorporated in various areas to perform tasks from network management to information management. The utilization of these agents in the wireless environment implies the need for the application to support disconnected mode. Mobile agents are programs that possess the ability to migrate among a network's hosts or to an arbitrary position of their choice.

Considering the characteristics and responsibilities of these agents, most frameworks are being built with the following tools:

- A specific programming language such as Java, C, C++, and Prolog
- Communication languages such as KQML, ACL
- Database languages like SQL, RDF
- String's representation such as XML, RDF



Table 1. Types of agents used in the studies mentioned in the literature review.

Agent Types														
Past Investigation s	User, Interface	Mediator	Wrapper	Store	Page preprocessing	Querying	Search, gathering, Collecting	Provide, Retrieval	Reasoning	Information generation	Information propagation	Information policym aking	Extract	Mobile
2016														√
Djamel 2011 Shakti														
Kundu	√			√			√	√	√					
2011														√
EFFANGA														V
2010 Shudong Zhang 2010 LUO	√	√	√	√	√									
2010 LUO Junwei	√						$\checkmark$	√					$\checkmark$	
2007 Yi Xiao	√			$\checkmark$		$\sqrt{}$	$\sqrt{}$		√					
2007 LUO Yingwei	√				√		$\sqrt{}$	√						
2007 Lizhen	√	√	√											
2007 Hua Hu	√						$\sqrt{}$			√	√			
2003 Dimitri	√													
2000 Travis								√						
Bauer								V						
1999 Yong								√						
S. Choi								•						

Analysing this table, it is noticeable that in most of the studies, researchers have developed User/Interface, Search/Information gathering/Information collecting or Provide/Retrieval agents in their works.

A study (Travis Bauer and David B. Leake, 2001) implemented that his framework was designed using the retrieval agents, Google bot and Alta bot. As discussed previously the multi-agent concept encourages each agent to make full use of their autonomy and intelligence to function collectively to proceed towards the common goal of the WebIE system.

### IV. COMPARISON BETWEEN IR SYSTEMS

The diversity of the Internet is characterized by open, complex, dynamic, and distributed properties that pave the way for the development of agent technology, and that agency technology offers many advantages over traditional methods. The agent itself is an autonomous computing entity that can find and utilize various information resources and services to solve problems and provide service for users independently. This section elaborates on the comparison between the discussed frameworks(Jian-Shuang Deng, Qi-Lun Zheng and Hong Peng, 2005).

Table 2. Analysis of Agent-based IR Systems

_			
System	Techniques/Area	Results	Challenges/ Drawbacks
(Choi and Yoo, 1999)	Artificial neural network, Back Propagation Neural Network, Yahoo! <i>Korea,</i>	Efficient Effective	Expense
(Hu and Xu, 2007)	XML, Java	Much faster, Dynamic Results, Better performance,	More storage, More low relevance rank data
(Kundu, 2011)		Meets user's demands	Privacy
(Liu <i>et al.</i> , 2007)	Resource Description Framework, JavaBean, XML	Uniform quick query, Retrieval service to users, Easy to integrate, flexible and extensible.	
(LUO Junwei, 2010)	Bayesian Probability Model, Support Vector Machine, Neural Network Algorithm	Obtain required information, used in knowledge & document management, search engine, and other applications that require searching large quantities of information to achieve the purpose of reusing and sharing information.	
(LUO Yingwei, 2003)	Java, XML	Interoperation among spatial information,	
(Menacer and Guenaoui, 2016)	SB-framework, Java, JADE SB-IR prototype, XML	Mobility capability	Lack better relevancy
(Shudong Zhang, Ye Qin and Naiming Yao, 2010)	URL KB, DOM, XML, ACL	Act according to web page changes, Information extraction, Relying on sample learning, Multi-slot-based extraction rules Pattern match technology	Adaptability Expansibility
(Xiao, Xiao and Zhang, 2007)	OWL, XML, RDF, SQL, JAVA	Solve the bottleneck	Construction Maintenance User privacy

However, the implementation done by the past researchers shows that each of these frameworks includes at least one unique characteristic. Most of the investigations were done by XML, Java, and RDF (Resource Description Framework) result can be drawn upon further observation.

## V. CONCLUSION

Based on these analyses, it can be concluded that a sequence of research on Agent-based IR systems that have been conducted in the past two decades has brought about new insights related to multi-Agent systems. The studies done around the semantic web as well as intelligent agents,



along with other different technologies such as information retrieval, knowledge modelling and ontology construction through years have led to the development of agents-based intelligent retrieval frameworks in the semantic web. The Traditional information retrieval methods are directly influenced and managed by the user, whereas newer and more dynamic methods of retrieving information are prompted by data inputs and events. For traditional information retrieval, relevant data for different criterion are extracted on demand and does not require to be stored, though for dynamic information retrieval setting and storage of such data will take place in advance. As they work on-demand, the results given by Traditional information retrieval methods are both static and retrieved once at a time. Meanwhile, the latest techniques to retrieve information along with their many dynamic updates, bear the ability to give results that are visibly evolving. Some data that are of lesser relevance can be collected with dynamic information retrieval techniques, resulting in low information veracity. The rules should be set and stored in advance in dynamic information retrieval. The need for temporary results to be restored gives rise to the need for more space in dynamic retrieval, in the system.

Since relevance and ranking in dynamic information retrieval only handle the changes occurring in the information, the effort exerted in computing can be cut down significantly, which results, in much faster, accurate and effective information retrieval compared to traditional retrieval methods.

#### REFERENCES

Bargain (1999) 'Bargain Bot'. Available at: http://www.ece.curtin.ed.au/%60sanub/bargainbot.

Choi, Y. S. and Yoo, S. I. (1999) 'Multi-agent learning approach to WWW information retrieval using neural network', in *Proceedings of the 4th international conference on Intelligent user interfaces - IUI '99*. New York, New York, USA: ACM Press, pp. 23–30. doi: 10.1145/291080.291086.

EFFANGA et al. (2011) 'INFORMATION RETRIEVAL USING JADE MOBILE AGENT SYSTEM', World Journal of Applied Science and Technology, 3(1), pp. 106–111.

Hu, H. and Xu, B. (2007) 'An Agent-based Framework for Intelligent and Dynamic Business Information Retrieval', in *Workshop on Intelligent Information Technology Application (IITA 2007)*. IEEE, pp. 77–81. doi: 10.1109/IITA.2007.16.

International Telecommunication Union (2015).

Available at: https://en.wikipedia.org/wiki/International\_Telecommunication Union.

Jian-Shuang Deng, Qi-Lun Zheng and Hong Peng (2005) 'Information retrieval from large number of Web sites', in *2005 International Conference on Machine Learning and Cybernetics*. IEEE, pp. 2172-2177 Vol. 4. doi: 10.1109/ICMLC.2005.1527305.

Kundu, S. (2011) 'Design of web data mining in agents based e-commerce', in 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011). IEEE, pp. 1–4. doi: 10.1109/ICCCT.2011.6075208.

Liu, L. et al. (2007) 'Cooperative Work for Agent-Based Heterogeneous Information Integrated Retrieval in Digital Libraries', in 2007 11th International Conference on Computer Supported Cooperative Work in Design. IEEE, pp. 395–399. doi: 10.1109/CSCWD.2007.4281468.

LUO Junwei, X. X. (2010) 'Research on Information Retrieval System Based on Semantic Web and Multi-Agent', in *2010 International Conference on Intelligent Computing and Cognitive Informatics*. IEEE, pp. iv–iv. doi: 10.1109/icicci.2010.3.

LUO Yingwei, W. X. and X. Z. (2003) 'Spatial Semantic Network and Agent-based Framework for Spatial Information Interoperation', in *The National Science Foundation of China under Grant No.60203002: the National High Technology Development*, pp. 40–47. doi: 10.1109/IRI.2003.1251393.

Menacer, D. E. and Guenaoui, N. (2016) 'A Market Collaborative Agents-Based Framework for Information Retrieval', in 2016 International Conference on Collaboration Technologies and Systems (CTS). IEEE, pp. 465–471. doi: 10.1109/CTS.2016.0088.

Prasad Kantamneni, R. G. and Narayanan, S. (2001) 'Personalization of information retrieval through user profiling', in 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236). IEEE, pp. 3475–3478. doi: 10.1109/ICSMC.2001.972058.

Rajendran, T. and Balasubramanie, P. (2009) 'An efficient framework for agent-based quality driven Web Services discovery', in *2009 International Conference on Intelligent Agent & Multi-Agent Systems*. IEEE, pp. 1–2. doi: 10.1109/IAMA.2009.5228044.

Shudong Zhang, Ye Qin and Naiming Yao (2010) 'An agent-based system framework for multi-slot Web information extraction', in 2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010). IEEE, pp. 200–203. doi: 10.1109/CAR.2010.5456664.

Travis Bauer and David B. Leake (2001) 'Handling Complex Information Environments: A Multi-Agent Framework', pp. 68–71. Available at: http://ceurws.org/Vol-59/6Bauer.pdf.



Xiao, Y., Xiao, M. and Zhang, F. (2007) 'Agents-Based Intelligent Retrieval Framework for the Semantic Web', in *2007 International Conference on Wireless Communications, Networking and Mobile Computing.* IEEE, pp. 5352–5355. doi: 10.1109/WICOM.2007.1312.

#### ACKNOWLEDGMENT

The researchers of this review extend our sincere gratitude to those who helped in the successful completion of this work in numerous ways. We thank General Sir John Kotelawala Defence University for providing us with this opportunity to showcase our efforts and improve our research potential.

We are also most grateful to Dr. Budditha Hettige for guiding and encouraging us in this research towards its completion.

### **AUTHOR BIOGRAPHIES**



Gayamini Gnanasuganthan working as a Lecturer (Prob.) in the Department of Computer Engineering, Faculty of Computing, General Sir John Kotelawala Defence University. This author was awarded the

M.Sc. degree in Computer Science Technology from Wuhan University Technology, Wuhan, P.R. China in the year 2020. She received her B.Sc. (Joint Major) degree in Electronics and Computing & Information Systems from the Wayamba University of Sri Lanka in the year 2017 and has excelled in the field of academics lecturing in subjects related to Electronics and Computer Science. Her current research interests lie in Artificial Intelligence, Multi-agent systems, Machine Learning, IoT and Robotics & Automation.



Naduni Thamalka Jayathilake, currently reading for her M.Sc. in Applied Electronics at the University of Colombo, obtained a B.Sc. (Hons.) in Computer Engineering from

General Sir John Kotelawala Defence University, along with a B.Sc. in Electronics and Automation Technologies from the University of Colombo. Her fields of interest include multi-agent systems, sensor networks, IoT and wireless networking.