Cognitive Ontology based Framework for Networking Women in Sciences

Afolabi I.T., Oladipupo O.O. and Ekanem I.E.

Department of Computer and Information sciences, Covenant University, Ota, Nigeria
{ibukun.fatudimu, funke.oladipupo,idongesit.ekanem}@@covenantuniversity.edu.ng

Abstract:

In order to increase the percentage of women in academics or researchers, there is need for a functioning research networking through which women can exchange ideas; ask questions and more importantly, mentorship, in Nigeria. In order to make for this, several recommendations have been suggested but are not scientific. Therefore, to bridge this gap scientifically, this paper is presenting an overview of a question and answering system framework that hybridize semantic search methodology and cognitive reasoning. The hybridization will enhance the question and answering accuracy especially due to the introduction of domain ontology for the semantic process. This paper also presents the output of the first phase of the implementation which is the development of the domain ontology for the question and answering system.

Keywords: *Ontology, ontology-based information extraction, Question and Answer System, Cognitive,Women,Science*

1 .0 Introduction

Recent report, reveals that only 12% of third year female PhD students want a career in academia (Curt , 2012). According to (Chandler, 1993; Newsome, 2008), the results of a longitudinal study with PhD students in chemistry in the UK reveal that, 72% of women express an intention to pursue careers as researchers, either in industry or academia but by third year it dropped to 37%. It was also discovered that women find academic careers as all-consuming, solitary and unnecessarily competitive. In Nigeria, according to a descriptive survey carried out in University of Lagos, with hardly any exception, the global picture is that men out numbering women at about five to one at middle management level and at about twenty to one at senior management level, especially in science and technology faculties (Udeani, 2011). As solution to the above problem, many recommendations have been made by various researchers. Akinsanya, (2013) recommended that Women achievements should be celebrated to encourage other women function well. Also in Bird, (2003) mentorship was suggested to enable female scholars to better understand higher education and their place within the academy. In order to attack this problem scientifically this paper therefore introduces an hybrid of cognitive model for the semantic search and an ontology based question and answering system to facilitate mentorship of upcoming female academics. This is based on question and answering system which understands the intent of users question and introduces the concept of ontology into the semantic search process. This gives adequate answers to questions based on the intent of the user.

The rest of the paper is structured as follows: In Section 2 we give the related work, section 3 is the overview of the proposed framework, Section 4 is report of work in progress, i.e. the development of the domain ontology for women in sciences and section five is the conclusion.

2.0 Related Work

Search engines have its relevance based largely in statistics i.e. counting links and words with little or no consideration to semantics. Introducing semantics presents difficult problems and the solution is more than bivalent logic and probability theory (Zadeh, 2003). Question-answering system (Q/A) is a system with deduction capability that can synthesize an answer to a question by drawing on the knowledge base (Zadeh, 2003). In Rohini (2000) an information extraction question and answering system that is based on Named Entity tagging was discussed and for future work a correlated entity that would help to provide better answers was recommended. PiQASso is a QA system that select paragraph that contains answer to requested questions. It performs semantic analysis of retrieved paragraph by checking the presence of entities of the expected answer type and extracting logical relations between words. But in doing the above, the system does not handle unknown names (Attardi et. al., 2003). Also in 2002, Deepak explored the power of surface text patterns to answer questions. Patterns are automatically extracted from documents and the precision of each pattern is calculated. The patterns are then used to find answers to new questions. While this method works well for finding date of birth for example it has problems with question types such as definition (Deepak, 2002).

(Zheng, 2002) developed a system named AnswerBus which is an open-domain question answering system that is based on sentence level Web information retrieval. It determines potential answers by classifying all words two categories of matching words or not matching words. This system was reported to outperform similar systems when evaluated. The IBM research division highlighted the current problems of QA systems. These include accuracy, (percentage of questions for what the QA provides answers) confidence accuracy, the probability that the QA is right, Broad domains, question difficulty, query complexity and so on (Ferrucci et al, 2009).

Introducing semantics into QA systems brought about ontology based QA systems. As mentioned by (Lopez, 2005).Many systems simply use ontology as a mechanism to support query expansion in information retrieval, but AquaLog provides answers derived from semantic annotations to queries expressed in NL. It is also implemented on-line and has a wider linguistic coverage (Lopez, 2005). (Guo and Zhang, 2008) presents a QA system that represents domain knowledge as ontology, it is based on three models, question's semantic comprehension model based on Ontology and Semantic Web, FAQ-based question similarity match model and document warehouse-base automatic answer fetching model. Though, it has been proven to be a feasible method but it is yet to be used to develop a Question Answering System. In (Lopez, 2007), PowerAqua provides the first comprehensive attempt at supporting open domain QA on the Semantic Web. The work is still ongoing on extending the range of queries the system is able to handle, include more ontologies and provide a better user interface. Furthermore, the system performs automated question and answering by converting natural language question into ontology based query by using an algorithm based on free text to ontology based search. It is recommended that the system's performance be enhanced by improvements to the term expansion (Serhatli, 2009). In (Palaniappan, 2010), a new method for ontology-based question answering (QA) where hypothesis questions and query templates are produced from domain ontology was introduced. Different QA systems were compared in (Kalaivani, 2012) and the following limitations were discovered; the quality of the annotation depended on the quality of the ontology and this is costly because its domain dependent. There is also lack of appropriate reasoning and so on. Kalaivani, 2012, proposed a semantic search methodology which uses Graph Matching Algorithm for query matching with the ontology using Spread Activation Algorithm. This approach could be improved with the use of semantic search model instead of using normal keyword search model. With this new research development in question and answering system this paper provides an overview of a framework that is an hybrid of semantic search methodology and the cognitive reasoning as a recommendation for networking women in science. (Kalaivani, 2012; Wechsler, 2004)

3.0 Overview of the Approach

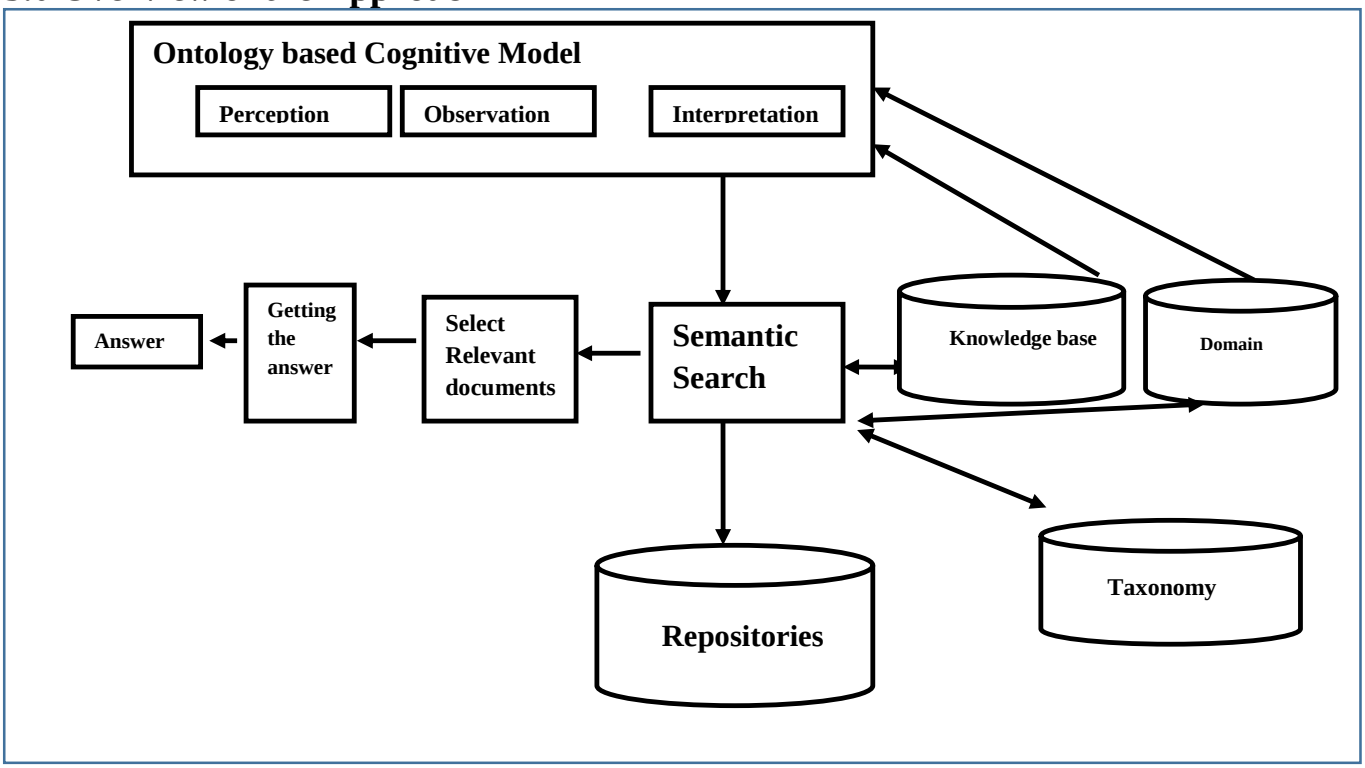

Figure 1: Framework for a Cognitive, Semantic, Question Answer System.

Perception

From Figure 1 above, the very first step in the cognitive model is perception and it aims at determining the motivation of the user of the system i.e. what the user is asking. The algorithm receives the question breaks it down to the list of words so as to know the ontology class associated with them. The ontology is combination of a general knowledge ontology and domain specific ontology created for the application area.

Observation Level

The goal of this stage is to describe the perception given above. The user describes his interest and any additional background information that may be useful in a paragraph in natural language. This information is used to list classes or individuals of ontology that is of interest to the user.

Interpretation

The next step is to find the ontology subgraph that is semantically closer to both the keyword and the elements in the profile. This resulting subgraph corresponds to the area of interest of the user or to the context that the user gives to the words.

Semantic Search:

This stage receives as input, the output of the interpretation stage ie the area of interest of the user. This output is used is used to select relevant document, thereby reducing the search space in the repository. To select the relevant document, the extracted words in the question gotten from the perception stage are taken as a word format and the relevant concept is searched in ontology and knowledge base. The Search will be carried out using Conceptual Graph Matching algorithm (Kalaivani, 2012). All the sentences in repository are framed as conceptual graph and the given question is also framed as conceptual graph. The matching of question CG with given CG are checked out using CG matching algorithms and the result us displayed at front-end of the QA system(Kalaivani, 2012).

Repositories and Ontologies

These repositories contain all the documents related to this Domain ontology field. The subject domain ontology is created semi-automatically from textual information sources of the domain such as publications and curriculum vatae.

Getting answer: This is a simple pattern matching technique to choose the appropriate response in terms of accuracy and simplicity.

Answer: After ranking, the browser displays the answer in the text field. If the user needs more information regarding it, the query will be given to the server once again.

There are various approaches to semantic search, but we chose this because of the application domain, it is particularly suited for mentorship purpose, because it takes in the profile, thereby providing some background knowledge of the person asking the question.

4.0 Description of the Domain Ontology

An ontology is used to define a common vocabulary for researchers to share information in a domain (Natalya & McGuinness, 2001). The importance of the domain ontology in the framework described in Figure 1 includes the following; It will serve as a standard vocabulary, which will foster a shared understanding of key concepts of the domain, it will enable effective reuse of domain knowledge and create a platform to formalize scientific information such that it becomes useful for man and systems. Due to this importance, the domain ontology development therefore forms the first phase of the implementation. This ontology developed is described in section 4.1 to 4.3.

4.1 Ontology Requirements

The ontology developed for Women in Science represents information of women in science academia using Covenant University as case study. This ontology is built such that the main user of the ontology is the question and answering application, as described in the framework in Figure 1. There also exists the secondary user of the ontology who is the prospect. The prospect is a user seeking for mentorship and therefore interacts with the ontology through the question and answering system. Example of these types of users is the female science students. The last category of users is the Domain expert. This is the person responsible for maintaining the concepts and relationships in the Women in Science ontology. The task of this user includes matching of concepts to existing ontologies and augmenting the ontology according to the new updates to the domain.

The ontology should be able to answer the following competency questions for example?

Who is Afolabi?

What is Afolabi's research trend?

What is the research relationship between Afolabi and Ayeipola?

What is the connection between Afolabi and Biological Sciences?

4.2. Ontology Modelling and Development

The core idea of the approach to developing the Women in Science ontology is based on (Noy and McGuinness, 2000) and outlined as follows.

(i) Determine the scope of the ontology by answering questions such as, which domain are we interested in? What will the ontology be used for? Who will use the ontology etc. The next step is to decide the source for the ontology i.e domain experts, documents or existing ontology.

(ii)The next step is to select relevant textual information using information extraction process extract all the keyword and key phrases that exist in them.

(iii) Build the ontology using the following steps; Enumerate important terms, define the concept taxonomies, relations, attributes, instances, axioms and functions Create an application interface through which the ontology can be accessed for the purpose of querying.

The first step has been detailed in section 4.1. The second step is the Information Extraction stage. It includes the selection of the key terms in the relevant domain documents. For this case study, the main source of the documents is the CVs of the female faculties in the college of Science in Covenant University downloaded from the websites.

Relevant terms were selected from the document corpus using the information extraction process which includes the following; filtration and stemming. The main concepts and properties are them selected from the extracted keywords and organized into concepts and properties hierarchically into categories;

To implement the ontology, the protégé-based development approach by (Noy and McGuinness, 2000) was adopted. Also to demonstrate the competency of the ontology, an interface for the prospective user was developed using the Java netbeans and Jena API in order to query the ontology.

The approach used to developed the class hierarchy is a top-down development process starts with the definition of the most general concepts in the domain and subsequent specialization of the concepts (Noy and McGuinness, 2000).

Main Concepts

The main concepts in women in science ontology is based on the need of the main actors, the questions and answering system and the prospect.

The following is description of few of them;

Person: The design of the concept *Person* is captures human beings in the domain academic domain chosen. Some of the individuals in this concept includes *Afolabi, Oladipupo, Usikalu, Ayepola, Okuboyejo.*

Field of study: This is used to capture the area of specialization of the concept *Person*. Examples include *Computer and Information Sciences, Microbiology* etc, it has the following sub concepts, *Agriculture, Business, Engineering, Environmental Studies, Human Resources, Medicine, Science, Social Sciences.*

Institution: This is used to capture the type of institution the concept *Person* has a form of relationship with. It has the following sub concepts, *University, College of Education, Polytechnic, Secondary School and Primary School.* Some instances includes, *Covenant Unversity, Babcock University, All Saints Church School.*

Position: Position refers to the academic status of the concepts *Person*. The individuals of this concept includes *Assistant lecturer, Professor* etc.

Publication: Publication refers to the type of out let that a particular research finding is published in. It has the following sub concepts, *journals, conferences, book chapter* and *book*. Instances includes *IJKBO (International Journal of Knowledge based Organizations), ICDDM, IBIMA* etc

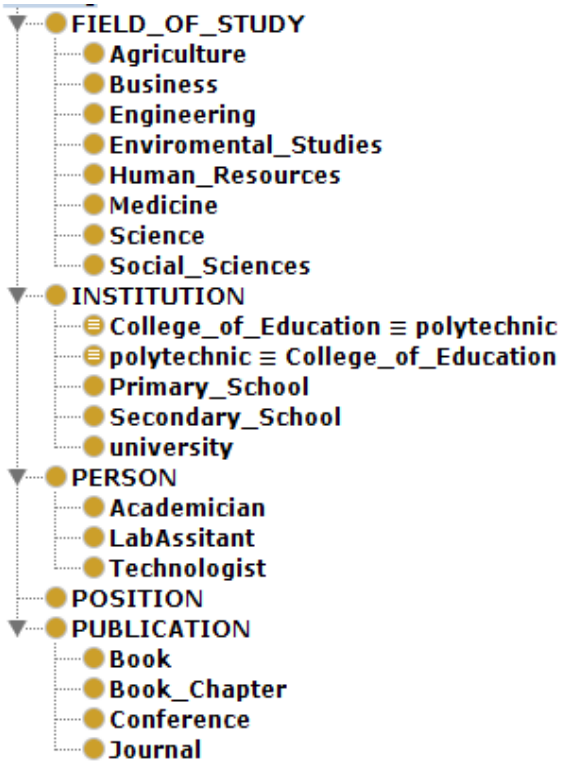

Figure 2: Main Concepts

Figure 2 is the protégé implementation of the main concept in the Women in Science Ontology. The Owlviz of the concept described above is presented in Figure 3.

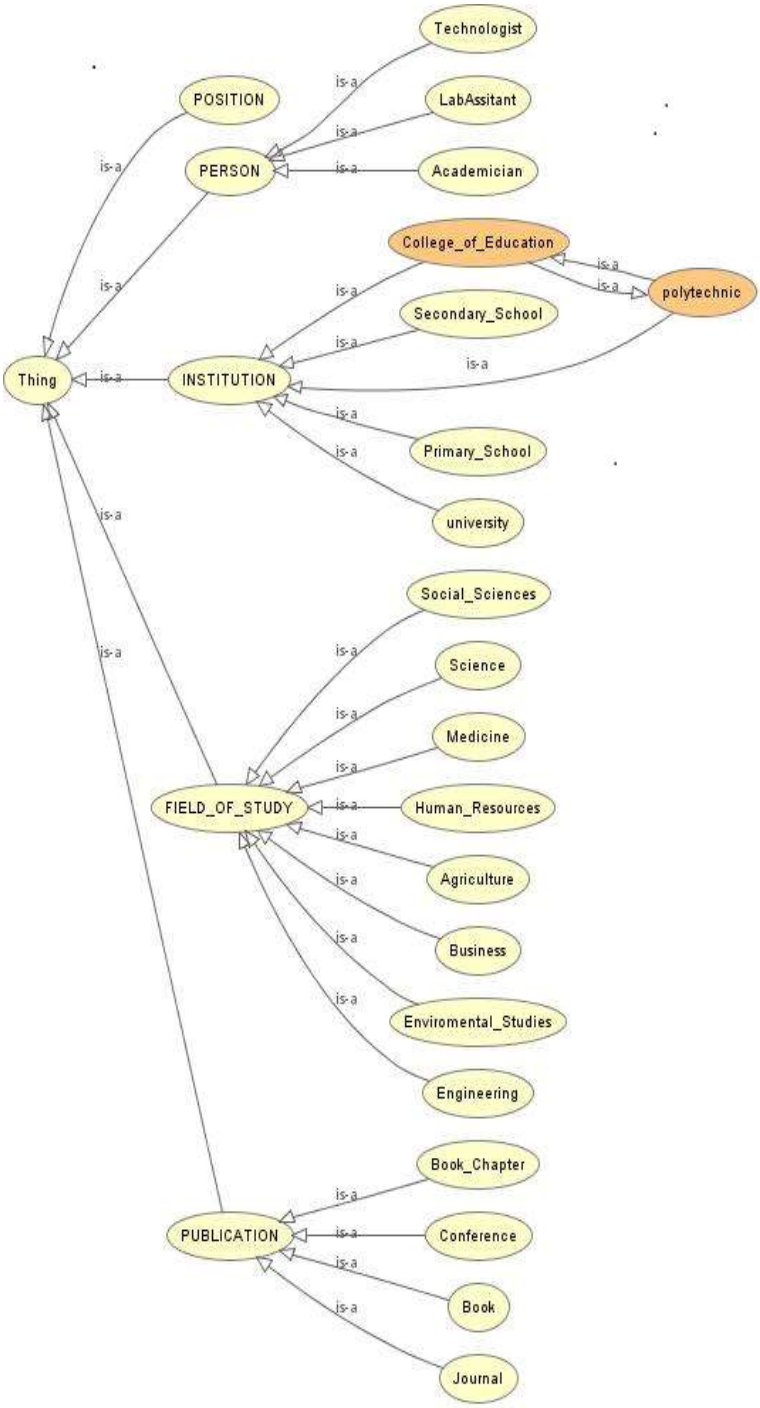

Figure 3: OwlViz presentation of the main concepts.

Exploring Ontology Constructs

This section is a description of the main constructs used to represent the women in Science Ontology. When exported to RDF(S) ontology is codified using the following constructs.

***rdfs**:label*: Used to store the general or popularly known name used to refer to the concept or individual. For example the individual IJKBO is labelled *International Journal of Knowledge based Organizations*

rdfs:DefinedBy: Used to store the definition of the term. For example the concept *Science* can be defined as *"*Science is a systematic enterprise that builds and organizes knowledge the form of testable explanations and predictions about the universe".

rdfs: Comment: Captures the further Explanation on the term or individual. For example, the individual *Oladipupo* can be further explained as *"Date of Birth: August 26th 1974, Place of Birth: Modakeke, State: Osun State Nationality: Nigerian, Marital Status: Married*

Residential: Flat 3, Block Z2 New Estate, Covenant University, Ota. Telephone Number:08033319370"

owl:equivalentClass: Used to equate similar concepts in meaning for example Polytechnic is equated to *college of Education*.

owl:sameAs:Used to equate similar individuals.

***rdfs:subClassOf* :** Used to break down general concepts to the type of classes that make them up.

The Women in Science ontology contains the following OWL Object properties, *Attended, CollaboratesWith, GraduatedFrom, IsA, SpecilizesIn, WorksFor.*

4.3 Ontology Competence

An application interface is developed for the Women in Science ontology in order to evaluate the competence of the ontology. This is developed for the Prospect i.e, the user seeking for mentorship as a female in the Science domain. The application is designed used the Java programming language (Net beans IDE). It give the prospect opportunity of searching for mentorrs. The Onotology was developed using the Protégé Owl whcin is connected to the Java application interface using the Jena API. The ontology is then queried using the Sparql query.

Querying the onotolgy from this application interface for response to the sample competency question, *Who is Oladipupo?*, the user is able to search for the term revealing the output in Figure 4.

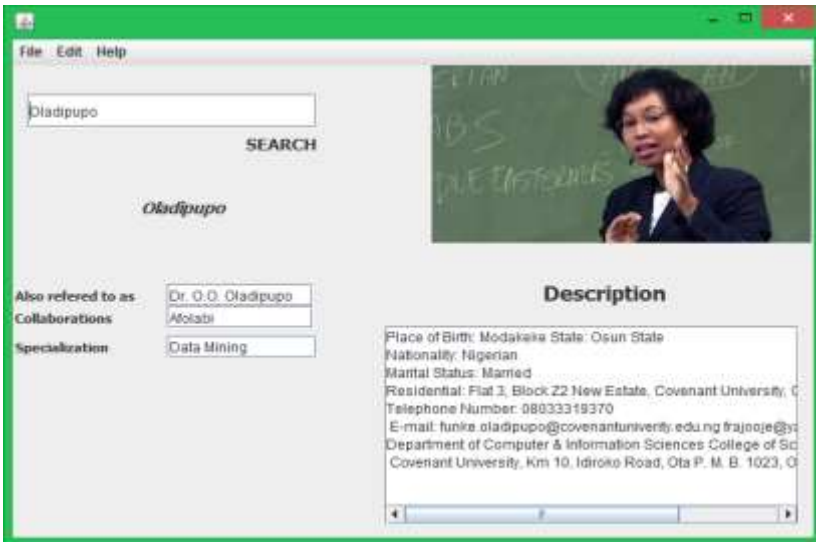

Figure 4: Out displaying the ontology competence.

The interface also allows search using a substring in case the user is not able to remember the exact term being searched for.

The Women in Science Ontology contains 25 concepts categorized into 5 top-level concepts. The maximum depth of the concept hierarchy is 2. Table 1 presents some metrics on the Women in Science Ontology. All concepts and individuals have definition and further explanation where necessary. Also of note is that the development of ontology is an on-going process, in order words there is provision for the ontology to be updated through the platform of the Web 2.0.

Table 1 Women in Science Ontology Metrics

Metrics	Values
Number of concepts	25
Number of ObjectProp	6
Subconcept axioms	20
Equivalent concepts axioms	1
Object property range axioms	6
Object property domain axioms	6
Number of individuals	761

5.0. Conclusion and Further work

This paper is a report of work in progress of a cognitive ontology based framework for networking women in sciences. On completion, this framework will facilitate collaboration of female academic researches across various universities and research centres. The approach proposes to provide edge over normal search engines because of the semantic processing capabilities. Even though this system does not seek to eliminate the input of other researchers who are not females but it seeks to help upcoming female researchers have direction in research, feel comfortable in a chosen field, and have female mentorship.

REFERENCES

1. Akinsanya O. O. (2012)"The Role of Women in Academics: Issues, Challenges And Perspectives" in JORIND 10 (2), June, 2012. ISSN 1596-8308Pp 136-141
2. Attardi G., Cisternino A., Formica F., Simi M., Tommasi A. (2003), "PIQASso 2002", in Proceedingsof The Eleventh Text Retrieval Conference (TREC 2002), NIST Special Publication SP 500-251, 2003.
3. **Bird** K. S. (2013) "Do women academics really publish less than men?" *The Guardian*, 22 January[Online] Available at http://www.guardian.co.uk/higher-education-network/blog/2013/jan/22/women-academics-publishing-ref-2014 (Accessed: 19 May 2013).
4. Chandler C. (1993) "Mentoring and Women in Academia: Re-evaluating the Traditional Model" Stanford University. Pp 79-100
5. Curt R. (2012) "Why women leave academia and why universities should be worried" *The Guardian*, 24 May[Online] Available at http://www.guardian.co.uk/higher-education-network/blog/2012/may/24/why-women-leave-academia (Accessed: 18 May 2013).
6. Deepak Ravichandran and Eduard Hovy(2002) "Learning Surface Text Patterns for a Question Answering System" *In Proceedings of the ACL Conference, 2002*

7. Ferrucci D., Nyberg E., Allan J. et al.(2009) Towards the Open Advancement of Question Answering Systems, IBM Research Report, RC24789 (W0904-093).
8. Guo Q.and M Zhang.(2008) "Question Answering System Based on Ontology and Semantic Web Rough Sets and Knowledge Technology Lecture Notes in Computer Science Volume 5009, 2008, pp 652-659
9. Kalaivani S. and Duraiswamy K. (2012) "Comparison of Question Answering Systems Based on Ontology and Semantic Web in Different Environment" Journal of Computer Science 8 (9): 1407-1413, 2012 ISSN 1549-3636
10. Kof, L. (2005). An Application of Natural Language Processing to Domain Modelling - Two Case Studies. *International Journal on Computer Systems Science Engineering 20*, (pp. 37–52).
11. Lopez, V., Motta, E., Uren, V. and Pasin, M. (2007). AquaLog: An ontology-driven Question Answering System for organizational Semantic intranets, *Journal of Web Semantics*, 5, 2, pp. 72-105, Elsevier
12. Natalya F. N. and McGuinness D. L. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001
13. Newsome J. L. (2008) "The chemistry PhD: the impact on women's retention" A report for the UK Resource Centre for Women in SET and the Royal Society of Chemistry Pp 1-38
14. Noy N. F. and McGuinness D. L.(2000)Ontology Development 101: A Guide to Creating Your First Ontology http://protege.stanford.edu/publications/ontology_development/ontology101.pdf
15. Omoronyia, I., Sindre, G., Stålhane, T., Biffl, S., Moser, T., & Sunindyo, W. (2010). A Domain Ontology Building Process for Guiding Requirements Elicitation. In R. Wieringa and A. Persson (Eds.): REFSQ 2010, LNCS 6182, (pp.188–202).Springer-Verlag Berlin Heidelberg.
16. Palaniappan L. and Rao S (2010). "An Ontology-based Question Answering Method with the use of Query Template" *International Journal of Computer Applications (0975 – 8887) Volume 9– No.9, Pp 23 -27*
17. Serhatli M. and AlpaslanF. N. (2009) "An Ontology Based Question Answering System on Software Test Document Domain" World Academy of Science, Engineering and Technology, PP 654- 658
18. Srihari, R. and W. Li. 2000. A Question Answering System Supported by Information Extraction. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics* (ANLP-NAACL-00), 166–172.
19. Udeani U. (2011) "A DECADE INTO THE 21ST CENTURY: NIGERIAN WOMEN SCIENTISTS AND ENGINEERS HIGHLY UNDER REPRESENTED IN ACADEMIA" in The African Symposium (ISSN#TX 6-342-323) Volume11, No.2 Pp 99-105
20. Vanessa Lopez, Enrico Motta, Marta Sabou, Miriam Fernandez, (2007)PowerAqua: A Multi-Ontology Based Question Answering System – v1 *OpenKnowledge Deliverable D8.4 Pp1-14*
21. Wechsler K., Baier J., Nussbaum M., and Baeza-Yates R (2004). Semantic search in the www supported by a cognitive model. In *Advances in Web-Age Information Management:5th International Conference, WAIM 2004, Dalian, China, July 15-17*, volume 3129 of *Lecture Notes in Computer Science*, pages 315–324. Springer, 2004
22. Zadeh L.A. (2006) "From Search Engines to Question Answering Systems – The Problems of World Knowledge, Relevance, Deduction and Precisiation FUZZY LOGIC AND THE SEMANTIC WEB Edited by Elie Sanchez Pp 162 -210
23. Zheng, Zhiping. 2002. AnswerBus question answering system. In E. M. Voorhees and Lori P. Buckland, editors, *Proceedings of HLT Human Language Technology Conference (HLT 2002)*, San Diego, CA, March 24–27.