Design, Analysis and Implementation of Semantic Web Applications

Sani Danjuma[1,*], Jun Zhou[2], Hongyan Mei[3], Ahmed Aliyu[4] , Usman Waziri[5]
Department of Computer Science and Technology,
Liaoning University of Technology, Jinzhou
Jinzhou, China

Abstract: Semantic Search Engine (SSE) has been developed to improve Kumbotso Teacher Educator Journal (KUTEJ) search engine, the software architecture implement an applications in the Semantic Web, based on designs and implementation specified in the Semantic Search Method (SSM). This architecture relies heavily on the ontology that structure underlying data for the purpose of comprehensive and transportable machine understanding. Therefore, the success of the Semantic Web depends strongly on the proliferation of ontology. Ontology model has been used in defining the concept and the association about the SSM.

Keywords: semantic web, ontology, semantic search engine, Resource Description Framework, Web ontology.

I. Introduction

The Semantic Web is an extension of the current web in which information is given well defined meaning, better enabling computers and people to work in cooperation [1]. This development has also changed the way we think of computers. Originally they were used for computing numerical calculations. Currently their predominant use is for information processing, typical applications being database systems, text processing, and games. At present there is a transition of focus toward the view of computers as entry points to the information highways.

The objective in developing semantic web technology is to develop the standard of technology that can help computers to better understand web information, and to support semantic search, data integration, navigation and task automation etc [2]. These standards will facilitate us in carrying the following tasks:

i. More accurate return of results; producing more meaningful information when searching the information.
ii. Integration and comparison of information from different alloplasm sources.
iii. Correlation of meaningful descriptive information about certain resources.
iv. The attachment of detailed information on the web for the automation of web services

The traditional search engines are based on the information retrieval technologies and implement operations such as Boolean queries, proximity searches, text relevance and link analysis. Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers) [3]. The aim of an information retrieval system is to find relevant documents thus relevance is a (if not 'the') central concept of information retrieval [4]. Information retrieval can be identified as data retrieval, document retrieval, information retrieval, and text retrieval, but each also has its own body of literature, theory, praxis, and technologies [5].

Document retrieval is also known as Text Retrieval. Text retrieval is a branch of information retrieval where the information is stored primarily in the form of text. The document retrieval search engine works by matching text records (documents) against user queries. It consists of a database of documents, a classification algorithm to build a full text index, and a user interface to access the database.

In this paper, we present an approach for designing and implementing applications in the Semantic Web based on the Semantic Search Method. Section 2 presents a literature review section 3 discusses design and implementation architecture and section 4 draws some conclusions and points to future work.

II. Semantic Web Overview

The semantic web was introduced by Tim Bernes-Lee and Tim O'Reilly. Tim Bernes-Lee is the director of the World Wide Web Consortium (W3C). Tim Bernes-Lee defines that the semantic web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation [1].

The Semantic Web technology consist of the Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples), and notations such as RDF Schema (RDFS) and the Web Ontology Language (OWL). These technologies are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain [6]. Below is the hierarchical structure of semantic web:

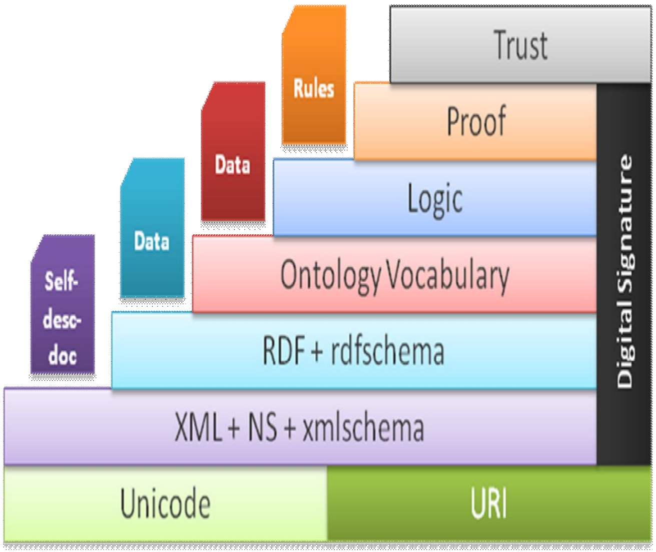

Figure 1: Hierarchical Structure of Semantic Web

Based on the figure 1: above, the lowest level of the hierarchical structure consists of a URI (Uniform Resource Identifier) and a Unicode. URI and Unicode address the method with which to access the resources of the web protocol. The next level contains the XML (eXtensible Markup Language) along with the namespace that can define this concept in modular terms. RDF (Resource Description Framework) and RDF schema, describing resources, are located on the next floor. Ontology is located on the fourth floor, and technological elements pertaining to law, logic, proof etc. are located in hierarchies above this [7].
.

2.1 Unified Resource Identifier (URI)

Unified Resource Identifier (URI) is a string or character that identifies resources or names on the internet. It is a way to identify any contents whether it would be documents, images, downloadable files, services, electronic mailboxes, and other resources. Uniform Resource Locator (URL) is the most common form of URI, which is a particular form or subset of URI called a Uniform Resource Locator (URL) [8]. Uniform Resource Name (URN) also can be used as a URI. URIs can be classified into two as locators (URLs), or as names (URNs), or as both. URN functions like a person's name, while a URL resembles that person's street address. In other words, the URN defines an item's identity, while the URL provides a method for finding it [9].

2.4 Resource Description Framework (RDF)

Originally RDF was designed as a metadata data model. It is a family of World Wide Web Consortium (W3C) specifications [10]. RDF is a standard model for data interchanged in the web. RDF has features that help data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. [11]. RDF is an application of XML(eXtensible Markup Language) and uses XML as a common syntax for the exchange and processing of metadata [5]. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.RDF provides a model for describing resources. Resources have properties (attributes or characteristics). RDF defines a resource as any object that is uniquely identifiable by a Uniform Resource Identifier (URI). The properties associated with resources are identified by property-types, and property-types have corresponding values. *Property-types* express the relationships of values associated with resources. In RDF, *values* may be atomic in nature (text strings, numbers, etc.) or other resources, which in turn may have their own properties. A collection of these properties that refers to the same resource is called a *description*. At the core of RDF is a syntax-independent model for representing resources and their corresponding descriptions. Figure 2: below illustrates a generic RDF description.

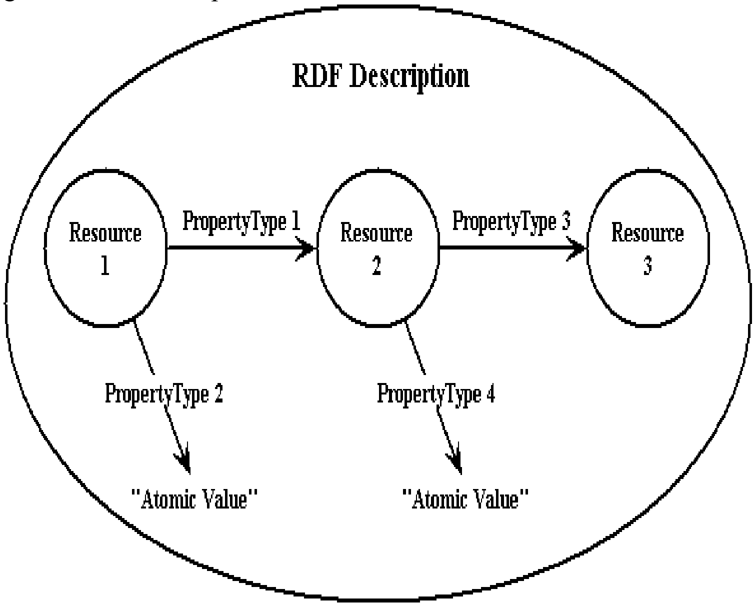

Figure 2: RDF Description

2.2 Ontology

Ontology is a description of concepts and relationships with a set of representational vocabulary The Semantic Web relies heavily on the formal ontology that structure underlying data for the purpose of comprehensive and transportable machine understanding. Therefore, the success of the Semantic Web depends strongly on the proliferation of ontology, which requires fast and easy engineering of ontology and avoidance of a knowledge acquisition bottleneck [12]. Ontology are (meta) data schemas, providing a controlled vocabulary of concepts, each with an explicitly defined and machine process able semantics. By defining shared and common domain theories, ontology helps both people and machines to communicate concisely, supporting the exchange of semantics and not only syntax.

2.3 Web Ontology Language (OWL)

The OWL Web Ontology Language is designed for use by applications that need to process the content of information instead of just presenting information to humans. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema (RDF-S) by providing additional vocabulary along with a formal semantics. OWL has three increasingly-expressive sublanguages: OWL Lite, OWL DL, and OWL Full [12].

2.4 Persistent Uniform Locators (PURLs)

PURLs are Web addresses that act as permanent identifiers in the face of a dynamic and changing Web infrastructure. A PURL provides a reference to an intermediate resolution service. Instead of resolving directly to Web resources, PURLs provide a level of indirection that allows the underlying Web addresses of resources to change over time without negatively affecting systems that depend on them. This capability provides continuity of references to network resources that may migrate from machine to machine for business, social or technical reasons. The PURL resolution service associates the PURL with the actual URL and returns that URL to the client. A PURL has three parts such as a protocol, a resolver address, and a name.

2.5 Semantic Web Applications

There are several search engines that has been developed using semantic web approach such as:

2.5.1 Swoogle

Swoogle interface. Swoogle is a search engine that used semantic web technology. It is a research project carried out by the ebiquity research group in the Computer Science and Electrical Engineering Department at the University of Maryland. Swoogle employs a system of crawlers to discover RDF documents and HTML documents with embedded RDF content. Swoogle has implemented main elements of the semantic web include data model description formats such as Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, Turtle, N-Triples), and notations such as RDF Schema (RDFS), the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms, and relationships within a given knowledge domain that is Wikipedia.

2.5.2 Hakia

Hakia is a semantic search engine and search structured corpora (text) like Wikipedia. The results of using Hakia search engine are organized in tabs: Web results, credible sites, images and news. All kinds of information on the subject can be found using this portal. Every results of keyword that we are searching for has an index of links to the information presented on the page for quick reference. The elements of these results will vary according to the nature of the query (e.g. biography, bibliography, timeline etc. for persons, government, economy, culture etc. for countries). The results are display by category [13].

2.5.3 Sensebot

SenseBot is a web search engine that summarizes search results into one concise digest on the topic of query. The search engine attempts to understand what the result pages are about. For this purpose it uses text mining to analyze Web pages and identify their key semantic concepts [14].

III. DESIGN STAGE

In the design process phase, use-case model has been used to design the process of KUTEJ Semantic Web Search Based on the requirement that has been gathered, two actors and four processes has been identified. The two actors consist of the user and the semantic web search engine. The four processes are Upload Paper, Define Ontology, Search Keyword and Display Related Article. Figure 3: shows the architectural model of KUTEJ Semantic Web Search.

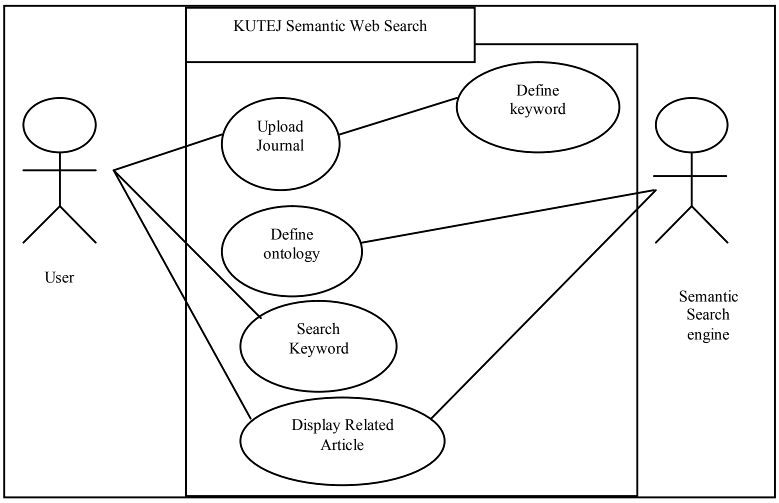

Figure 3: Model for KUTEJ Semantic Web Search

Ontology Model

All system for ontology learning that are able to construct useful ontology are semiautomatic[13],[15] in other word a considerable amount of human work is necessary to get the good results. Ontology model design is a vital part in designing KUTEJ Semantic Search Engine (SSE) application. Ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. Figure 4: shows the conceptual of ontology

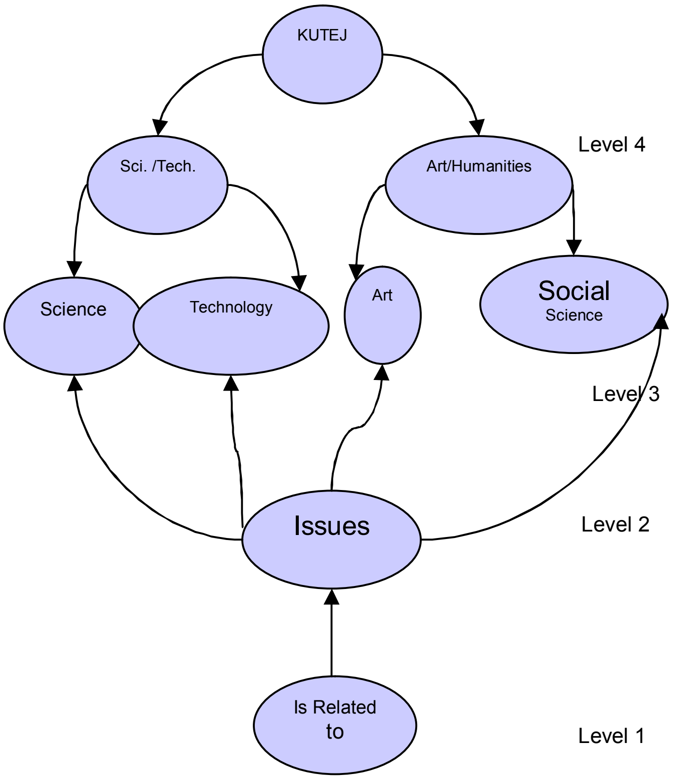

Figure 4: Ontology Model of KUTEJ SSE

IV. Survey

A survey has been conducted among lecturers, tutor and research assistants to evaluate the current KUTEJ Semantic Web Search to demonstrate the effectiveness of the KUTEJ Semantic Web Search. And compare it with the older version of KUTEJ site 40 lecturers have been selected randomly to participate in this survey.

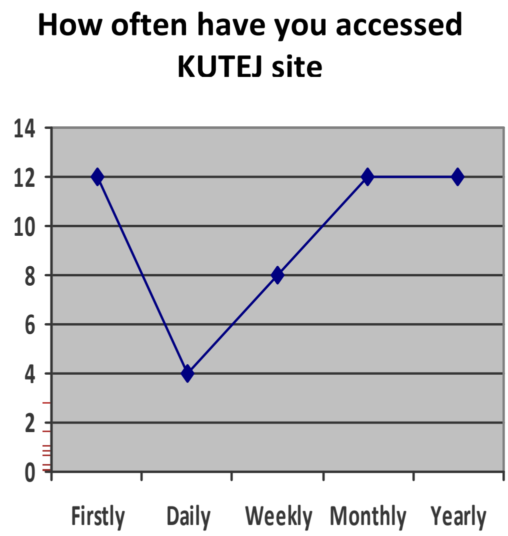

Figure 5: The frequency of user access to the KUTEJ Web

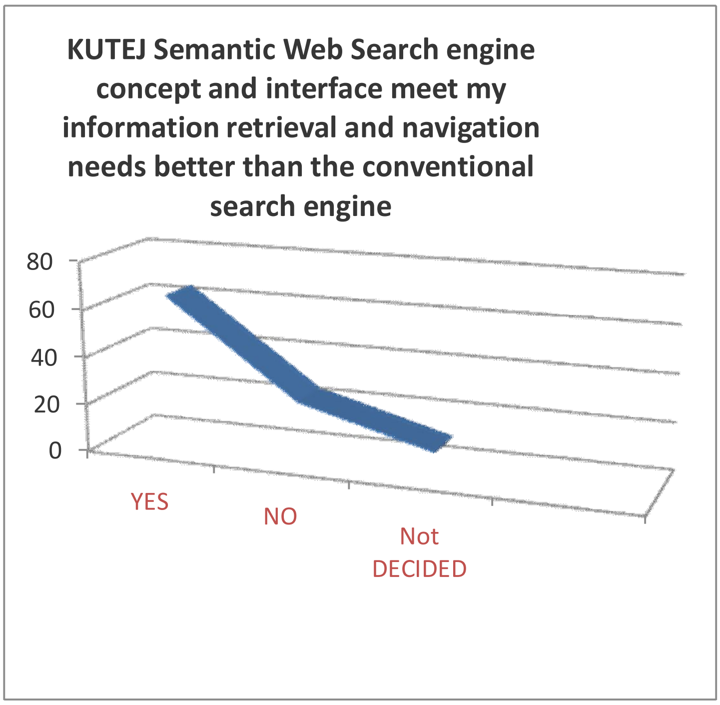

Figure 6: Number of lecturers uses KUTEJ Website

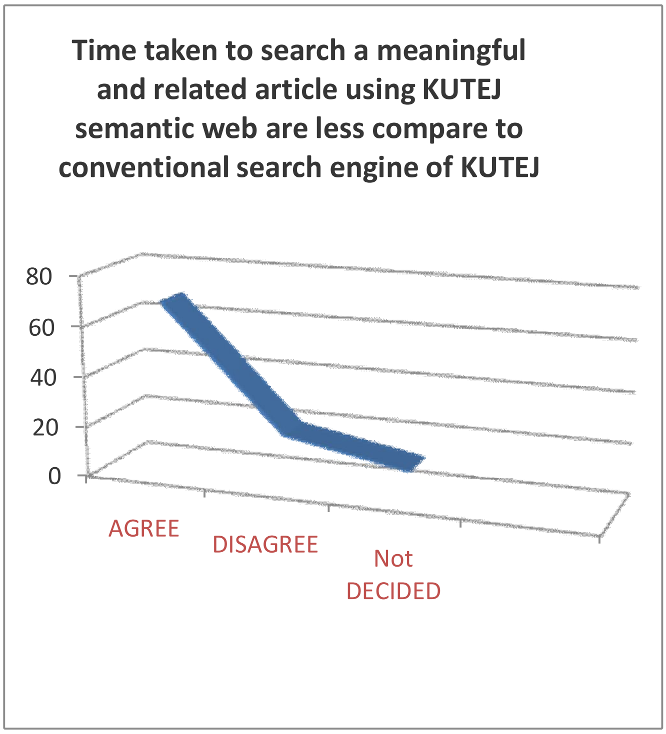

Figure 7: Number of lecturers agrees that time taken is less by using KUTEJ SSE

V. Discussion

From the results based on the system testing shows that using semantic web approach as a search engine has improve a lot in searching related and meaningful articles that related to the science, technology and humanities published by KUTEJ. The result shows that using the existing KUTEJ search engine, it returns many papers of articles and only few papers are meaningful and related to user. The difference is clear and it is quite big. But by using KUTEJ SSE, it returns relevant papers of articles and all the papers are meaningful and related to the user. The output results also display into groups and categories. This is very helpful to the user and will reduce the time taken to sift through the results return.

Based on the survey that has been conducted about 65% of user of KUTEJ web applications agree that using KUTEJ Semantic Web Search is better compare to the conventional search. Based on the results, the objectives of this paper have been met. The results show that by using the semantic web approach has improved a lot in searching related and meaningful articles.

VI. Conclusion and Future Work

As a conclusion, by implementing semantic web approach has improved the existing search engine by returning meaningful and related results. KUTEJ Semantic Web Search has made easier to the lecturers and other users in searching for a related article with a less time taken.

At the end further work is presented which is concentrated on the application and it is hope this work can be expanded by implementing searching an article from different web site. And also we cannot say how and to what extent the ontology can be used to build relation.

REFERENCES

[1]. Berners-Lee, T., Hendler, J., and Lassila, O., 2001. The semantic web. Scientific American, 284 (5), 34–43.

[2]. Nam-deok Cho and Eun-ser Lee. Design and Implementation of Semantic Web Search System Using Ontology and Anchor Text. Workshop on Approaches or Methods of Security Engineering (AMSE 2006, Sess. B). May 11, 2006

[3]. D. Manning, Christopher. Introduction to Information Retrieval. Cambridge university press, New York. 2008.

[4]. Mizzaro, Stefano (----). How many relevances information retrieval. University of Udine, Italy.

[5]. D-Lib Magazine, An Introduction to Resource Description Framework, May 1998. http://www.dlib.org/dlib/may98/miller/05miller.html. Retrieved 17th January, 2013.

[6]. Wikipedia, Semantic Web (2010). http://en.wikipedia.org/wiki/Semantic_Web. Retrieved 5th June 2010

[7]. Cranefield, S., 2001b. UML and the Semantic Web. Proceedings of the international semantic web working symposium, 30 July–1 August 2001, Stanford University, Stanford, CA. Stanford, CA: Stanford University.

[8]. Zhang, W.Y. and Yin, J.W., 2008a. Exploring SW technologies for ontology-based modelling in collaborative engineering design. International Journal of Advanced Manufacturing Technology,

[9]. Rumbaugh, J., Jacobson, I., and Booch, G., 1998. The unified modelling language reference manual. Reading, MA: Addison-Wesley.

[10]. Zhang, L., Yu, Y., Zhou, J., Lin, C., and Yang, Y., "An enhanced model for searching in semantic portals", In Proceedings of WWW '05.

[11]. W3C Semantic Web, Resource Description Framework (2010). http://www.w3.org/RDF/. Retrieved 6th February, 2013.

[12]. G. Stumme, A. Hotho, B. Berendt. Semantic Web Mining: State of the art and future directions. Journal of Web Semantics, Volume 4 (2), pp. 124-143 (2006)

[13]. Corcho, O.; Gomez-Pérez, A.; López-Cima, A.; López-García, V., Suárez-Figueroa, M-d-C; "ODESeW. Automatic Generation of Knowledge Portals for Intranets and Extranets", Proceedings ISWC 2003, LNCS 2870, Springer Verlag, October 2003.

[14]. Maedche, S. Staab. Ontology Learning for the Semantic Web. IEEE Intelligent Systems, Special Issue on the Semantic Web, 16(2) (2001)

[15]. G. Petasis, V. Karkaletsis, A. Krithara, G. Paliouras, C. D. Spyropoulos. Semiautomated ontology learning: the boemie approach. In: Proceedings of the First ESWC Workshop on Inductive Reasoning and Machine Learning on the Semantic Web, Heraklion, Greece (2009)

First Author Sani Dajuma: is working with Sa'adatu Rimi College of education, Kumbotso Kano state Nigeria. He did his first degree in computer science and currently, he is undergoing a master's degree program in Computer Science and Technology at Liaoning University of Technology, Jinzhou China. He is also a member of IEEE, he has written many papers among them is the paper titled "A critical Simulation of CPU Scheduling Algorithms using Exponential Distribution; An analysis for security in web applications; Using SQL for testing database application.

Second Author Jun Zhou: is the Dean School electrical/Electronic Engineering Liaoning University of Technology she did her first, second and doctorate degrees in computer science. Research area: Data mining, Pattern recognition, Artificial Intelligent

.
Third Author Hongyan Mei received her B.S. degree in computer application technology from China in June 1996 and her M.S. degree in computer science and technology from Jinzhou, china April 2008. She currently is working towards her Ph.D. degree in computer network services at Beijing University of Posts and Telecommunications, China. And she is also a teacher and works in Liaoning University of Technology. Her current research interest includes P2P network services and data mining.

Fourth Author Ahmed Aliyu is a Graduate Assistant at Department of Mathematical Sciences, Bauchi State University Gadau, Itas-Gadau Local Government, Bauchi State of Nigeria. And attended Abubuakar Tafawa Balewa University Bauchi State Nigeria, for his first degree in computer science and currently pursuing a master's degree in computer science at Liaoning University of technology, Jinzhou china.

.
Fifth Author: Usman Waziri Ahmad Aliyu is a Graduate Assistant at Department of Mathematical Sciences, Bauchi State University Gadau, Itas-Gadau Local Government, Bauchi State of Nigeria. And attended Bayero University, Kano, Nigeria, for his first degree in Mathematics and currently pursuing a master's degree in computer science at Liaoning University of technology, Jinzhou china.