Vol 8. No. 4 – December, 2015

African Journal of Computing & ICT

www.ajocict.net

CIDUCE: A Competitive Intelligence Tool for the Creation and Exploration of Document Usage

L. A. Akanbi*
Department of Computer Science and Engineering
Obafemi Awolowo University, Ile-Ife, Nigeria.
laakanbi@oauife.edu.ng

B. S. Afolabi
Department of Computer Science and Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria.
bafox@oauife.edu.ng

A. David
Université de Lorraine
Nancy, France.
amos.david@univ-lorraine.fr

E. R. Adagunodo
Department of Computer Science and Engineering
Obafemi Awolowo University
Ile-Ife, Nigeria.
eadagun@oauife.edu.ng

**Corresponding Author*

ABSTRACT

The amount of effort in terms of time and cognitive capacity required to browse through millions of available documents in order to identify useful documents that contribute significantly to the resolution of decision problems at any time are often enormous. This is due to the fact that documents are represented in the search systems with index which are terms derived from the document. These terms are often time orthogonal to what documents are eventually used for. In order to enhance accessibility to documents and reduce the amount of effort required to identify relevant and useful documents to the resolution of decision problems, a competitive intelligence based document usage creation and exploration system is developed. The architectural structure of the system is presented and detail description of various system processes (usage creation, usage exploration and usage based search operation) are done with the use of flowcharts. The system prototype is developed with the use of HTML for the front end and PHP scripting language and MySQL database management system for the backend. To evaluate the system, data was obtained through questionnaire and interview. Twenty postgraduate research students were asked to provide five documents that they considered useful in their research problem solving. The documents and research problems obtained were used to populate the database of the system. The result of system testing shows that the documents that would not have been considered as relevant to query terms were parts of the returned document based on previous usage.

Keywords: document, usage creation, usage exploration, competitive intelligence.

African Journal of Computing & ICT Reference Format:
L.A. Akanbi, B.S. Afolabi, A. David & E.R. Adagunode (2015): CIDUCE: A Competitive Intelligence Tool for the Creation and Exploration of Document Usage. Afr J Comp & ICTs. Vol 8, No. 3. Pp 119-134..

1. INTRODUCTION

Since the advent of the Internet and the evolution of web technologies, the amount of accessible information has continued to increase exponentially. Thanks to the advancement in web technology and the developments in information and communication technologies.

Baeza-Yates and Ribeiro-Neto in [4] rightly observed and reported in their book *Modern Information Retrieval* that "the web is becoming a universal repository of human knowledge and culture which has allowed unprecedented sharing of ideas and information in a scale never seen before.

Its success is based on the conception of a standard user interface which is always the same no matter what computational environment is used to run the interface. As a result, the user is shielded from details of communication protocols, machine location and operating systems".

The importance of information in day-to-day running of any organisation cannot be overemphasized. Menendez *et al*. in [12] pointed out that the flow of information is the lifeblood of organisations while the dramatic increase and acceleration in volume of information pose challenges for day-to-day management in all kinds of organisations. According to [18], the relationship between information and decision making is a complex domain which has been at the centre of research for several years, and more recently researchers have shown a relationship between information quality and quality in decision making which has consequences on the organisation strategy.

With the advancement in the Internet and web technology, the amount of information (documents) that usually match users' query has continued to increase exponentially. This access to millions of document sources has placed the onerous task of identifying relevant and useful documents that satisfy the users' demand at any time on the users themselves. The effort in terms of time and cognitive capacity required to browse through millions of available documents in order to identify useful documents that contribute significantly to the resolution of problems at any time are often enormous (for example it can take 2 – 3 minutes to flip through a document in order to determine whether it will be useful or not). This is due to the fact that documents are represented in the search systems with an index which contains terms derived from the document. These terms are often time orthogonal to what documents are eventually used for.

The existing techniques (e.g. [7], [11]) for integrating document context into document index are based on inference methods that use statistical or linguistic approach applied to the document index, but the limit of these methods is linked with the fact that the usage of information by the end-user cannot be automatically inferred from the index generated from the terms in the document. Therefore, to enhance accessibility to documents and reduce the amount of effort required to identify relevant and useful documents to the resolution of decision problems, a Competitive Intelligence based Document Usage Creation and Exploration (CIDUCE) system is developed and presented in this paper.

The rest of this paper is organized as follows. Related works are presented in section 2. In Section 3, the methodology adopted for the implementation of the system is presented. Section 4 describes the result of the system implementation and the conclusion is presented in section 5.

2. RELATED WORKS

The process of document selection for use in the resolution of decision problems depends on the user's cognitive ability, the prevailing situation and user's interest at the time. Wang and Soergel in [17] describe document selection as a decision process in which the user evaluates a retrieved document based on its surrogate obtained from a bibliographic information retrieval system to decide whether or not to further pursue the document, to browse the actual document, to obtain a copy of the document, to read the document, or to make a reference to the document. They identified three stages in the process of selecting document from a bibliographic search for use in decision problem resolution. Figure 1 shows the stages and decision points of document use along the stages of document seeking and use during a research project as identified by [17]. As shown from the figure, after obtaining the list of bibliographic citation through the information retrieval process, the first stage is selection. At this stage, the decision whether to pursue a particular document is made. The decision at this stage is based on the cognitive capability and experience of the user. At the reading stage, the user decides whether to read the documents or not. Those documents that are identified for reading enters the third stage. At the third stage, the user decides whether to cite the documents or not. The decision at this point is based on the contribution of the documents to the resolution of decision problem of the user.

According to [2], document is defined as a record or account of someone's experience or findings about a particular event. It could be composed of different objects such as texts, images as well as sounds. The objects in a document are usually arranged in an orderly manner in order that the information that the document is expected to hold and transmit is preserved and comprehensible.

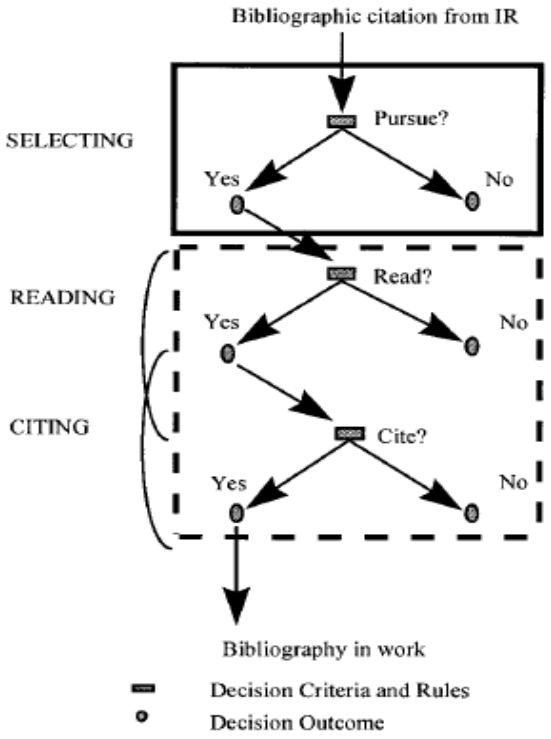

Figure 1: Document use at different Stage of Research Projects [17].

Competitive Intelligence (CI) is a special case of application of information retrieval process [3]. It is a systematic process of information gathering, processing, analysis and decomposition, which is conducted within the context of the external environment of an organization's activities. This process is with the major goal of supplying the right information, at the right moment, and in the correct structure, to the right person, in order to support the best decision possible [16]. The Strategic and Competitive Intelligence Professionals (SCIP) in [15] defines CI as the legal and ethical collection and analysis of information regarding the capabilities, vulnerabilities, and intentions of business competitors. The information legally and ethically collected is to be used by the decision makers in an organization to support them in arriving at the best possible decision. Dishman and Calof in [8] described CI as a process involving the gathering, analysing and communicating of environmental information to assist in strategic decision-making, and as such it is the fundamental basis of the strategic decision-making process. Camelo *et al.* [5] pointed out that the CI process should aim to provide information on technologies and general commercial tendencies in order to facilitate the decision making process, helping to achieve the strategic goal of the company.

According to Odumuyiwa in [13], CI is linked to various similar concepts such as economic intelligence, business intelligence and knowledge management and it is a process that embodies decision making. CI as an information search process is made up of the following phases according to in [6]: (i) identification of decision problem, (ii) transformation of decision problem to information search problem, (iii) identification of relevant sources of information, (iv) collection of relevant information, (v) analysis of collected information to extract indicators for decision making, (vi) interpretation of indicators and (vii) decision making.

Deerwester *et al*. in [7] addressed a fundamental problem that plagued the existing information retrieval techniques that try to match words of queries with words of documents. It was pointed out that users want to retrieve documents on the basis of conceptual content, but individual words provide unreliable evidence about the conceptual topic or meaning of a document. Singular Value Decomposition was used on the matrix of terms-documents formed from the related documents. The approach took advantage of implicit higher order structure in the association of terms with documents (semantic structure) in order to improve the detection of relevant documents on the basis of terms found in queries. Fur and Buckley in [9] identified some of the problems with the previous work on probabilistic indexing and presented a method for probabilistic document indexing that introduced the concept of "relevance description" as an abstraction from specific term-document relationships. The author reported that the approach showed improvement over other indexing methods.

Gupta *et al*. in [10] noted that the performance of a search engine is limited because of two problems, namely (i) Low precision, that is, irrelevance of many of the search results and (ii) Low recall, that is, the search engine is unable to index all the information in the web whereas un-indexed information may also be relevant. They presented a technique to index document efficiently by using hierarchical clustering to keep the information based upon similarity measure and fuzzy string matching. The method keeps the related documents in the same cluster so that searching of documents becomes more efficient in terms of time complexity. The work addressed the problem of information retrieval by indexing and putting related documents in the same cluster to enable easy access to information. Gupta and Sharma in their work on context-based indexing in search engines using ontology discovered that the current information retrieval systems use term to describe documents, and that the keyterm-based index seems to be less efficient due to two information retrieval problem, that is, polysemy (a word having multiple meaning) and synonymy (multiple words having the same meaning) [11]. The author proposed an indexing structure in which index is built on the basis of context of the document rather than on terms basis using ontology. The context of document was extracted using thesaurus and ontology repository.

A document usage model for CI process was developed by [2], in which documents usage was described based on the attribute value pair technique of document annotation proposed by [14]. A platform for integrating the document usage with the existing document index was developed in [3] where additional index (usage index) that captures what documents were used for, was created and integrated into the document index. The CIDUCE developed in this work was based on the ideas presented in [2] and [3].

3. METHODOLOGY

In this section, various techniques used in the development of the system are presented. The system architecture and structure are presented first followed by flowchart of various processes of the system.

3.1. CIDUCE System Structure

Figure 2 shows the structure of CI process with additional information on the stage at which the document usage model is integrated. After the first three stages of identification of Decision Problem (DP), transformation of DP into information search problem, and identification of relevant source of information, the next stage in the CI process is the collection of relevant information. Presently the information (documents) are represented with terms extracted during the process of document indexing.

As shown in the figure, the CIDUCE system makes provision for the representation of document with additional information (usage descriptors). Relevant documents based on the two representations of documents are returned as the result of the information collection stage of the CI process.
It is important to note that, to be able to determine the relevant sources of information to the resolution of DP, high level of cognitive ability and experience is required on the part of the user to see the connection between documents and DP. The effort of decision maker in identifying relevant documents to the resolution of DPs is what is hoped to preserve by the CIDUCE system, so that other decision makers can make use of the preserved knowledge in the resolution of similar DP in the future. This could also serve as learning tool for upcoming decision makers to get acquainted with DP resolution process in an establishment.

The flowchart of the CIDUCE system is presented in figure 3. As earlier discussed in [2], there are two categories of users (also known as actors) of the system. The first category is the decision maker and the second category is the information watcher. The decision makers in an organisation are the ones responsible for strategic decision making in the day-to-day running of the organisation. The information watcher is to support the decision makers in the process of decision making by providing necessary information when required. It is also the responsibility of the information watcher to keep the CI systems up-to-date. The first requirement to use the system is

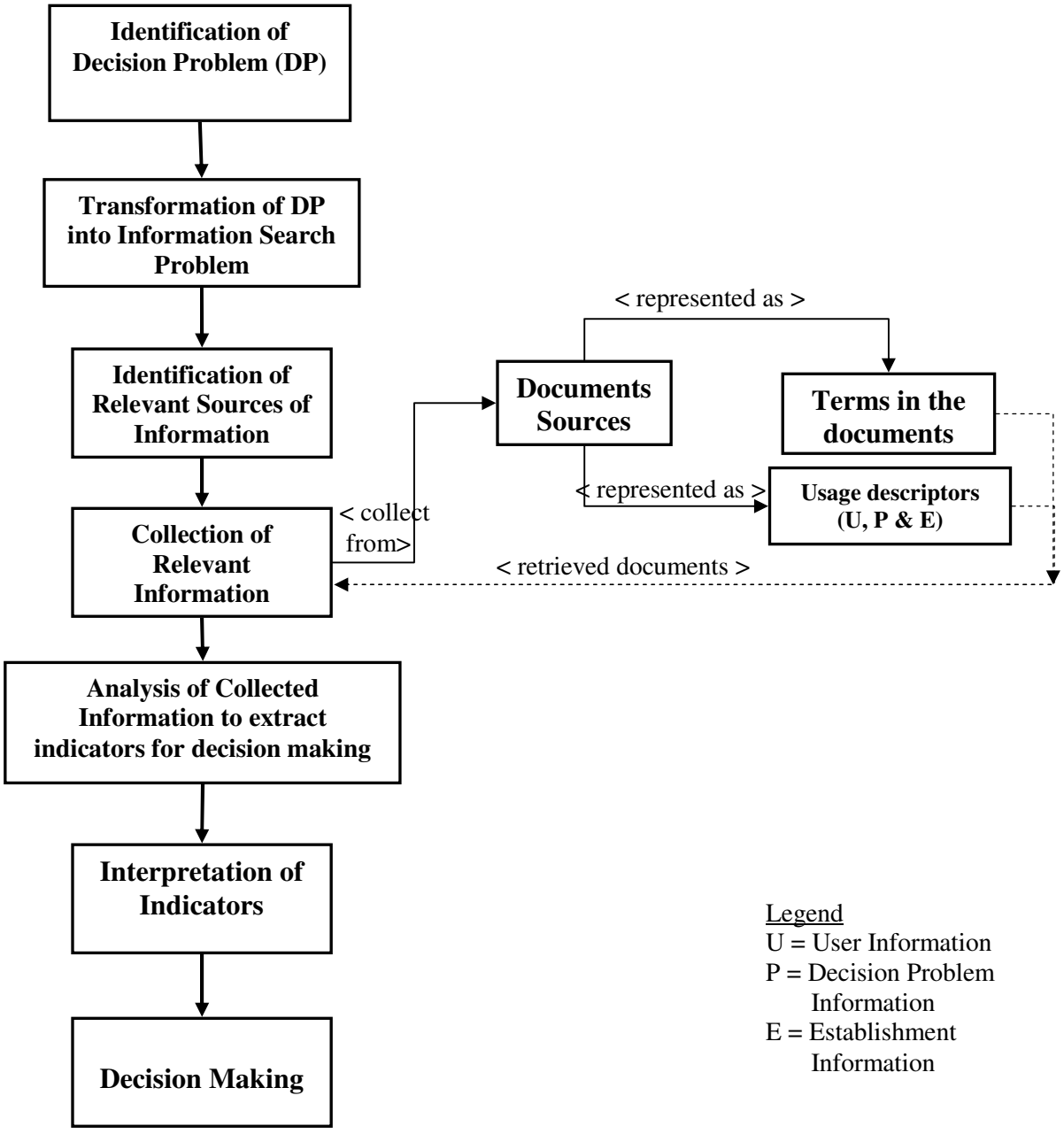

Figure 2: Structural Model of the CIDUCE System

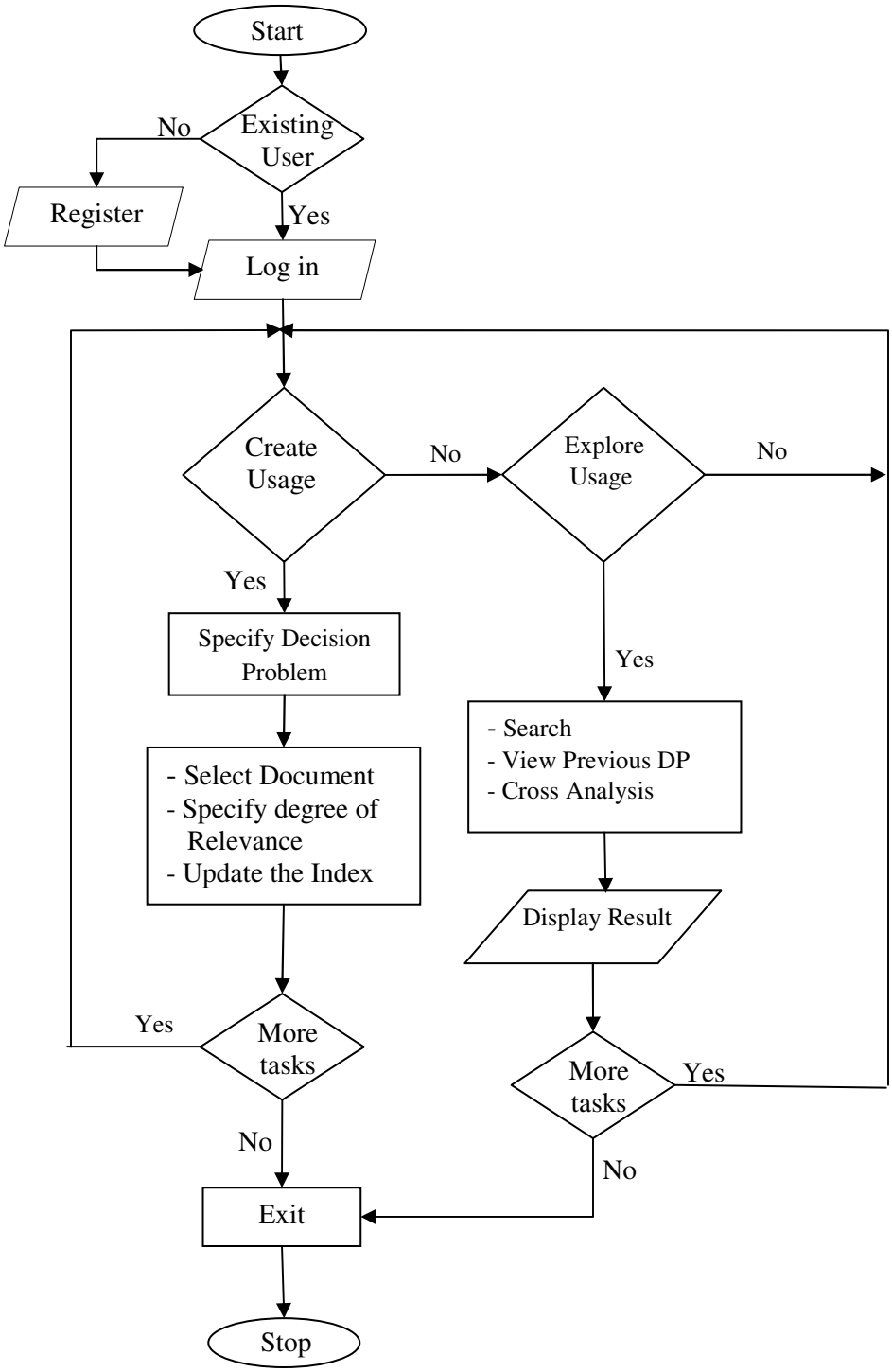

Figure 3: Flowchart of the CIDUCE System

registration. Any user, irrespective of the user category, must register to use the system. After successfully logging on to the system, the user can perform either of two functions/processes, namely i) Create usage, or ii) Explore usage. The usage creation summarily involves specification of DP, selection/identification of users, selection of documents for which usage would be created, specifying the degree of relevance of the document to the DP, and updating of the system. Within the usage exploration module, three activities (cross analysis of variables, browse through the previous decision problems and search through the system) are possible. The details are subsequently described in the following subsections.

3.2. Usage Creation Module

The usage creation module of the CIDUCE system is the module that implements the document usage model. Figure 4 shows the flow of activities necessary for the creation of document usage. In order to create usage for a document, the user after login on to the system must specify the DP by providing the stake object description for the DP. The category of user is then checked to determine whether it is the decision maker that is creating the usage directly or the usage is being created by the information watcher on behalf of the decision maker.

If it is the decision maker, then the system captures the logged user information and prompts the user to select the document of interest and specify the degree of relevance of the document to the resolution of the specified DP. Based on the DP earlier specified and document degree of relevance specified by the user, the document usage index is updated appropriately. If however, the user is an information watcher, then the system first prompts the information watcher to select the user on behalf of whom usage is being created. Other processes then follow in the same order stated earlier for the case of the user being decision maker.

3.2. Usage Exploration Module

There are three basic processes that can be carried out under usage exploration module. The usage can be explored by cross analysis of the variables (DP stakes, documents and users). The analysis allows the decision maker to see various relationships that may exist between the variables. From the analysis, the decision maker can know the number of documents that have been used during the course of resolving a DP, the number of DPs a document has been used solve, the number of users that had worked on a particular DP to mention but few. The result of the cross analysis can either be in textual form or in graphical (table and bar chart) form. Figure 5 is the flow chart diagram of the usage exploration process of the CIDUCE system. The second process is a search facility that allows the user to search for information on the system. Two mode of search exists within the facility. These are: search based on terms in the document, and search based on previous document usage.

The third process in the usage exploration module provides the ability to view previous DPs. This process also allows the decision maker to browse through the system to discover items that may assist him in the process of resolving DPs.

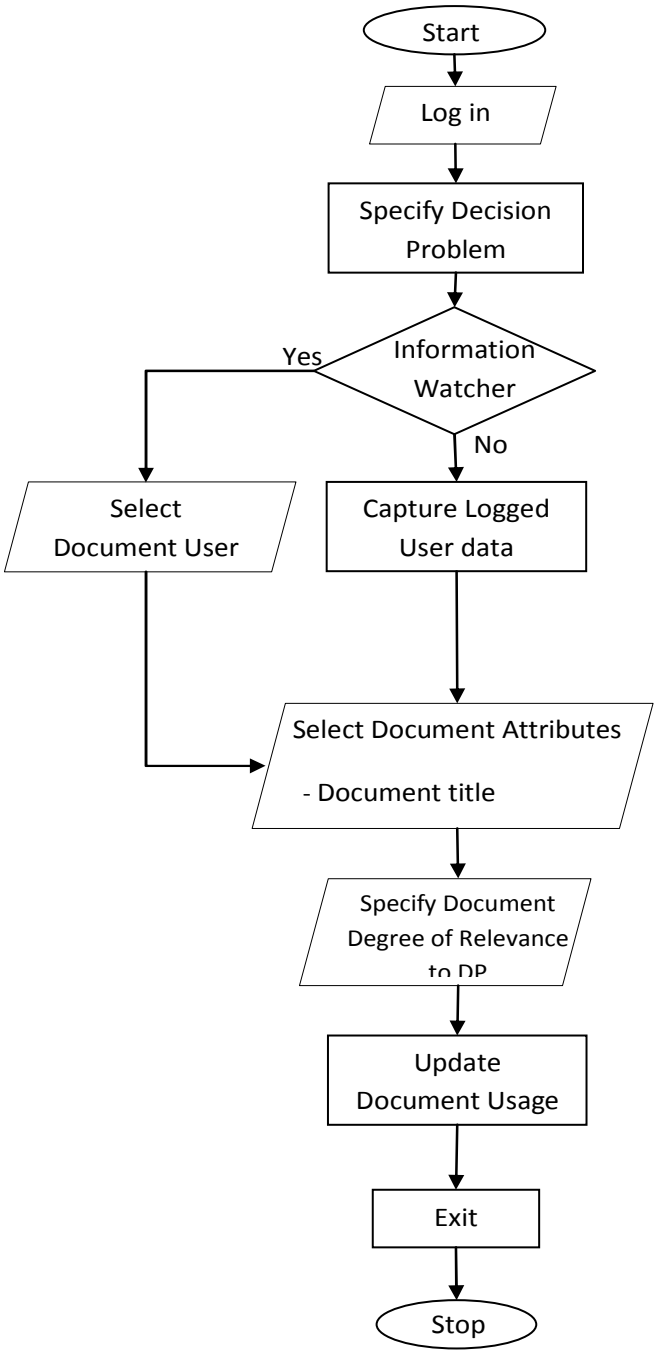

Figure 4: Flowchart of Document Usage Creation Process

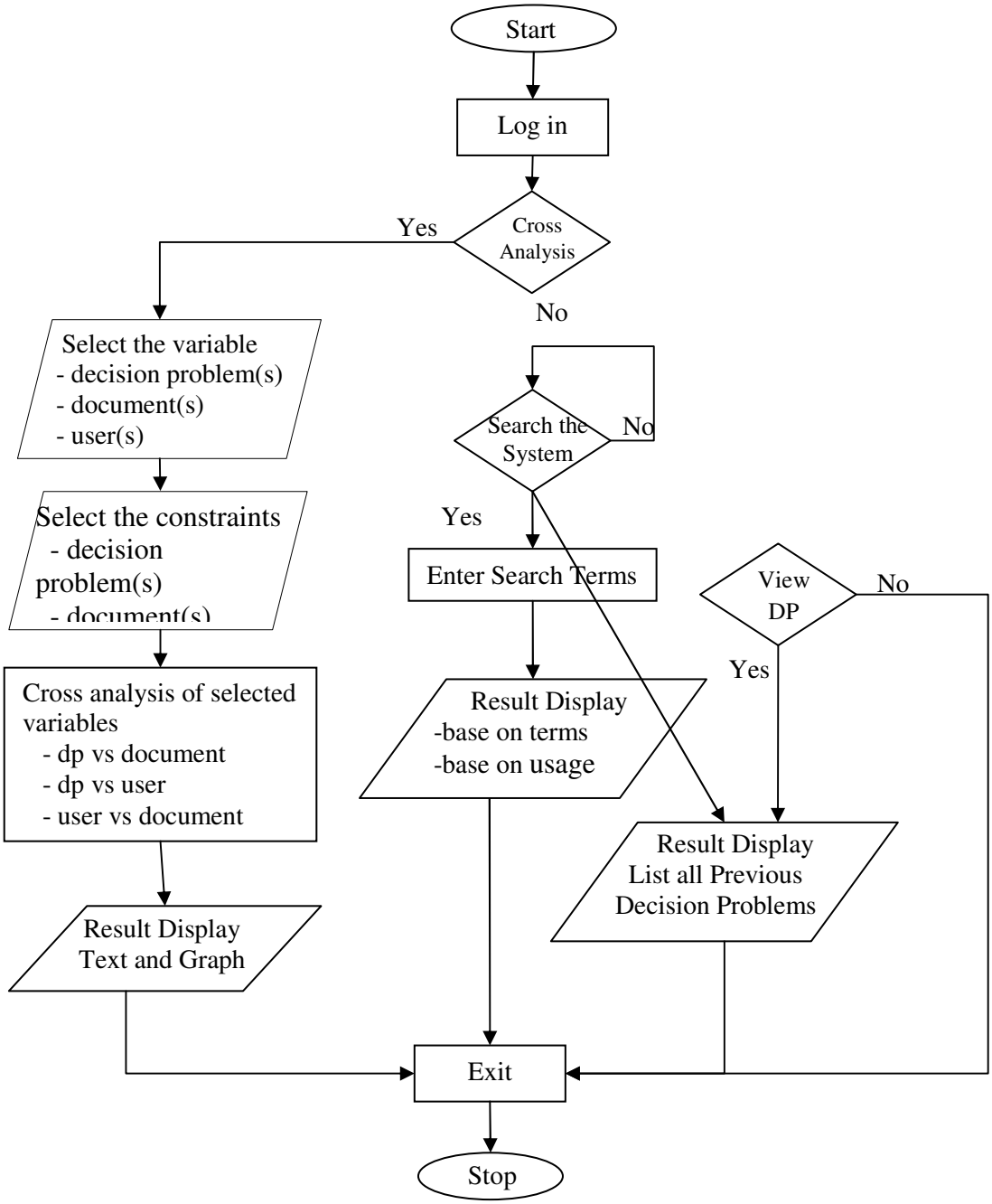

Figure 5: Flowchart of Document Usage Exploration Module

African Journal of Computing & ICT

www.ajocict.net

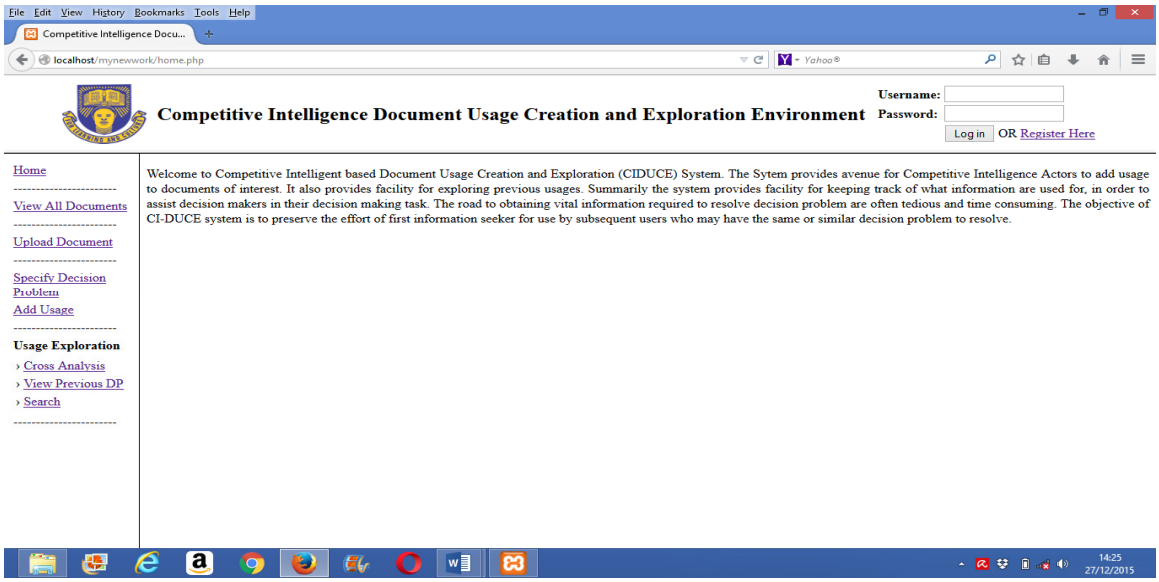

Figure 6: Home Page of CIDUCE System

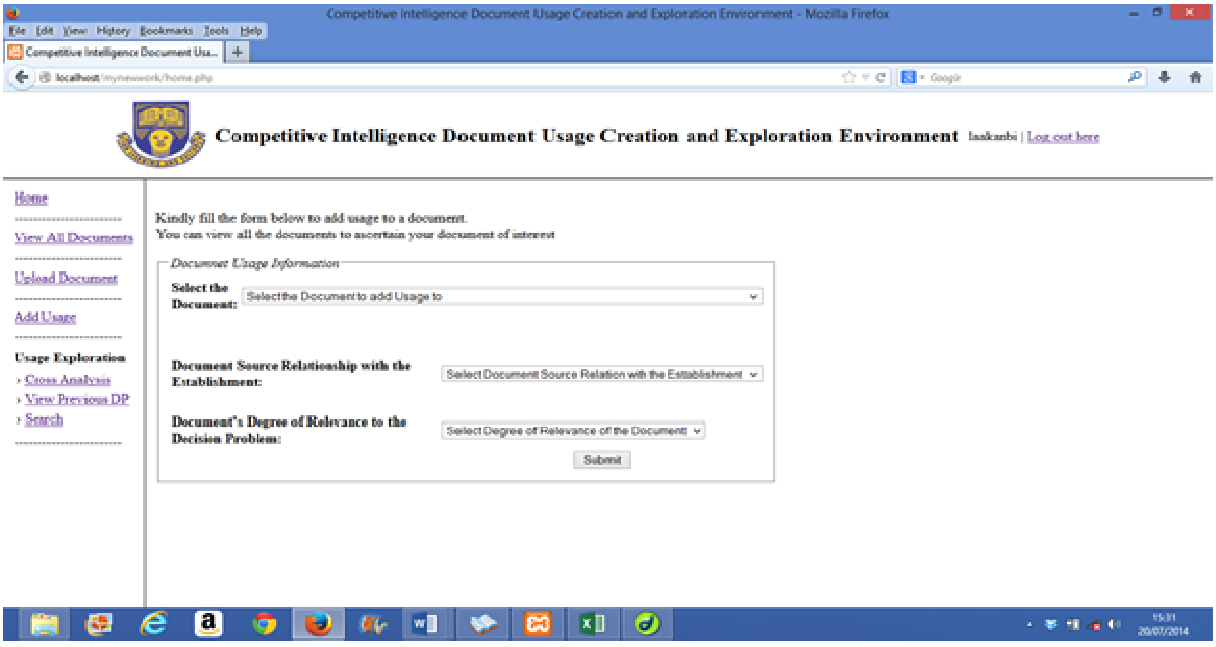

Figure 7: Interface for Adding Usage to a Document

3.3. System Implementation

The CIDUCE system is implemented with the use of HyperText Markup Language (HTML) for the front end design. The back end of the system is implemented with the use of Hypertex Preprocessor (PHP) and MySQL database management system. These programming tools were used because of their availability and ease of use.

4. RESULTS AND ANALYSIS

The home page of the CIDUCE system is shown in figure 6. The page contains the login input menu and it displays information about the system. The navigation bars on the left pane are used to interact with the system after logging on through the top pane. The right pane which is the biggest pane displays the result of the request issued by the users. The main functions of the system is grouped into two namely: (i) usage creation and (ii) usage exploration. To be able to create/add usage to a document of choice the user must provide information on the decision problem that had been solved with the document. After the specification of the DP, the user can issue add usage command to the system by clicking on the "Add Usage" icon on the left pane of the interface. The interface shown in figure 7 is presented to the user to select the document as well as its level of contribution of the document to the resolution of the DP. Information on the source of the document in relation to the establishment is also sought by the system through the interface. It should be noted that the system automatically harvests user information from the logged user information in the database.

Exploring the usage allows users to either carry out cross analysis to obtain information on previous usage of documents, or view previous problems solved in the establishment or search for documents. For cross analysis task, the user clicks the "Cross Analysis" icon on the left pane and the interface shown in figure 8 is presented to user. The user then selects the attributes as appropriate. Figure 9 is the result of the cross analysis in graphical form. From the figure, it is shown that 10 documents were used to solve one decision problem (the DP with id = 1 on the decision problem axis); while six, four, two and one documents were respectively used for the DPs with ids = 2, 4, 5, and 6.

Another important component of usage exploration is the search facility. Users can carry out two types of search, namely normal search and the usage based search.

The normal search provides search facility based on the terms in the document. This is the kind of search that is found in public search engines like Yahoo, Google, and Bing. The usage based search checks the document usage index for documents that match the users' query. Its results are based on previous usage of documents. Figure 10 is the interface for search operation on the system. As shown in the figure, the user should indicate the kind of search operation the system is expected to carry out by clicking the appropriate radio button.

As an example, figures 11 and 12 show the result of search for query term **"***grapheme to phoneme conversion***"** based on normal search mode and usage search mode respectively. The normal search operation returned a document that has query terms in its title, whereas the usage based search operation returned documents based on both terms from the documents and previous usage.

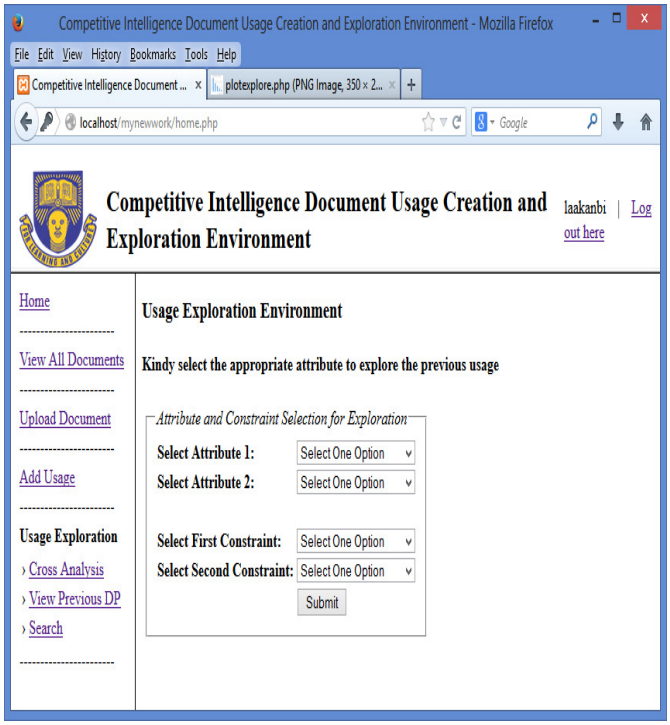

Figure 8: Interface for Cross Analysis Task

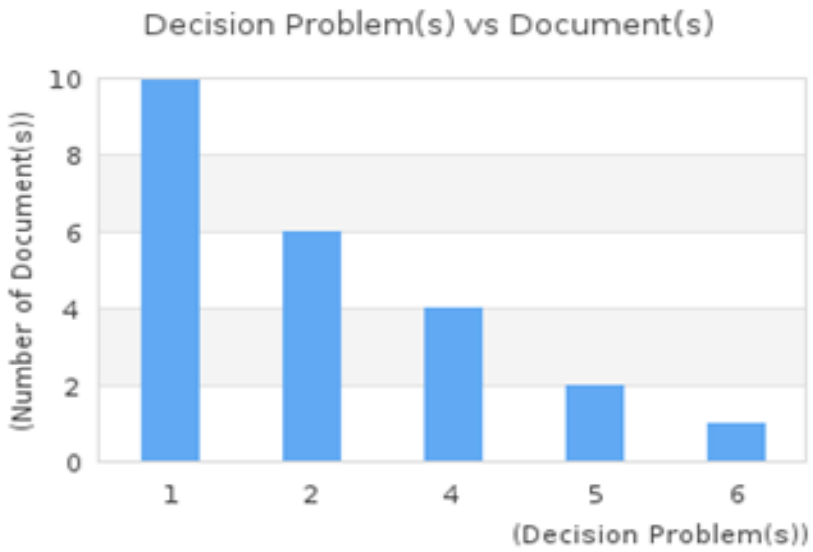

Figure 9: Graphical Representation of Cross Analysis Result

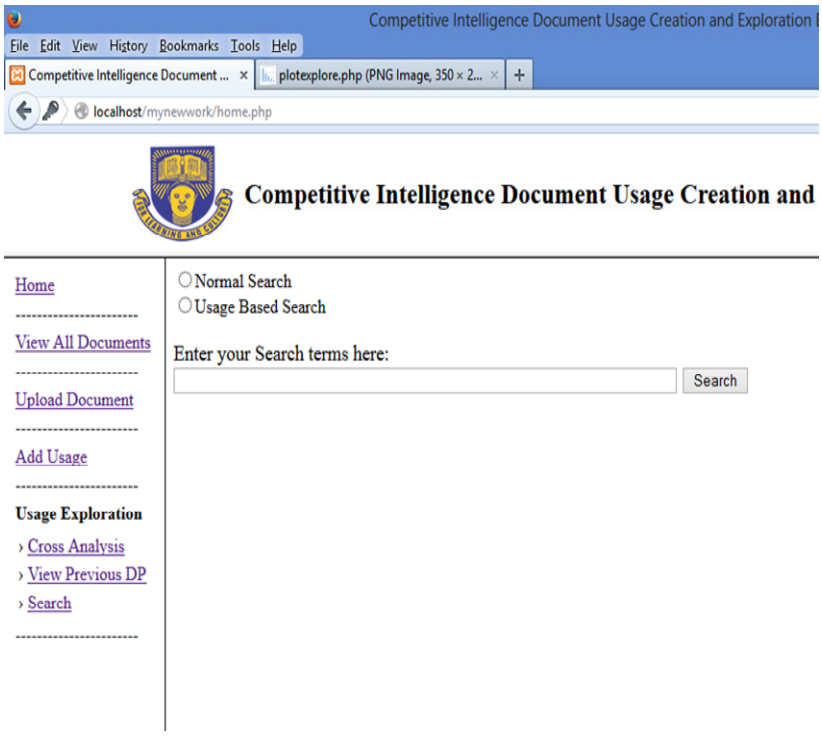

Figure 10: Interface for Search Operation

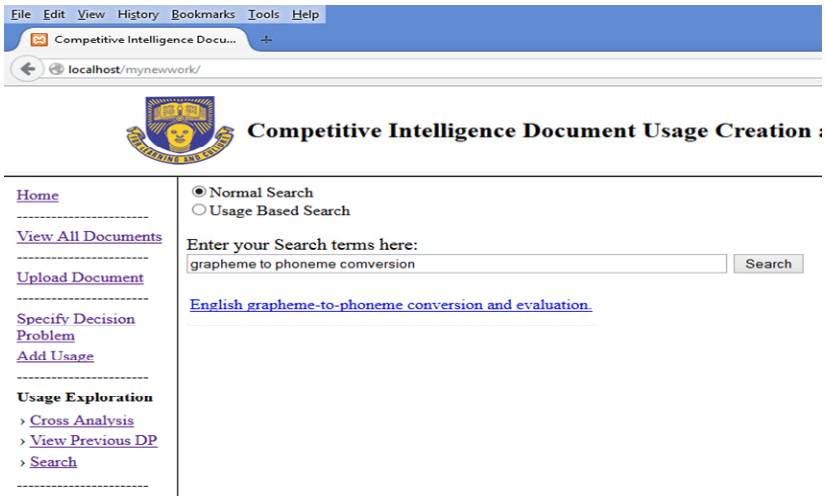

Figure 11: Result of Normal Search Operation

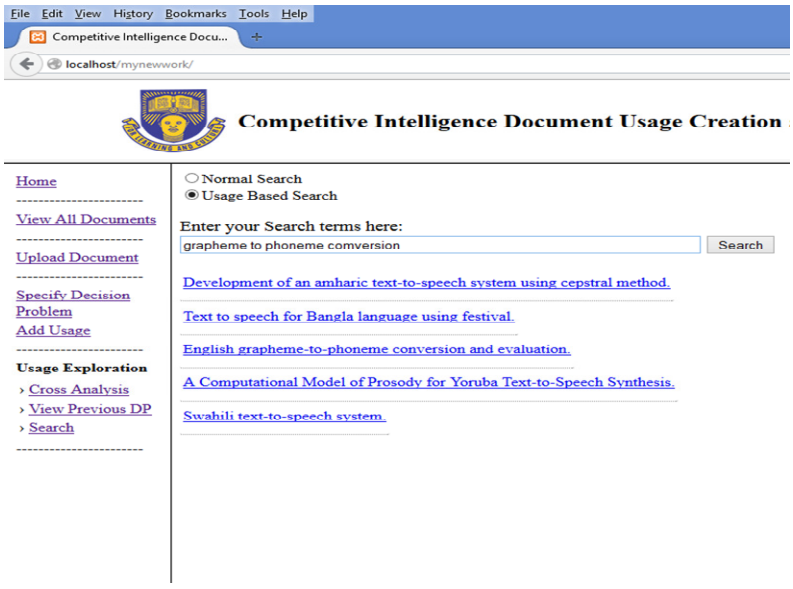

Figure 12: Result of Usage based Search Operation

The data used in the testing of the system was obtained through questionnaire administration and interview. Twenty postgraduate research students of the Obafemi Awolowo University Ile-Ife were asked to provide five documents that they considered useful in their research problem solving.

The documents and research problems obtained were used to populate the database of the system. The respondent were at the final stage in their research work.

5. CONCLUDING REMARKS

In this research work, a decision support system (CIDUCE) that will assist in preserving the efforts (cognitive-wise and material-wise) that users exert during the process of information search for the purpose of solving problems. The results from the evaluation of the document usage model established that, there is usually a gap between the terms in the document and terms in the decision problem of the user. This indicates that truly, users would have gotten the documents through the use of some other means. These means may be influenced by the experience, as well as cognitive capacity of the user. The study therefore concludes that in an information seeking process, there are usually documents in the document space whose index may not contain terms in the users query but which are very relevant to users' need. This work therefore provided alternative access to documents in the documents space, thereby increasing access to relevant documents and reducing the amount of efforts that would have been required to access them.

The idea presented in this work can find application in various industries that use the CI approach to resolve decision problems. Also the system can be used in the various units in the hospitals. For example, medical doctors can use the system to keep track of the process of treating patients (from diagnosis to drug prescription). The system can later be used by another doctor that have similar case at hand, it could also be used to examine how each doctor is faring in the hospital. A pharmacy unit of the hospital could use the system to keep track of the kind of medical conditions each drug is used to address. This will enhance the quality of decisions to be made on each drug when there is need for such.

Another application area is in the legal business (bar and bench). At the bar level, each law firm can employ the use of the system to keep track of the documents and cases they have been used to resolve. The document usage history can be used to train upcoming lawyers as well as assisting experience lawyers in getting access to relevant documents. At the bench level, the Courts or the judges can use the system to keep track of documents (exhibits and previous judgements) that influence their decisions on cases. The system could therefore serve as tool for training upcoming judges.

6. CONTRIBUTION TO KNOWLEDGE

The modest contribution of this work is in the area of establishing that documents relevance in terms of usefulness in the resolution of decision problems is strongly dependent on the ability of the users to formulate query terms that will lead to relevant documents. The work also provided a framework for creating document usage index that could be part of document search space during information search process.

7. FUTURE WORKS

The work is hoped to be extended to document usage cluster generation in order to enhance the quality of information search space.

REFERENCES

[1] Akanbi L. A., Afolabi, B. S., David, A. and Adagunodo, E. R. (2014) Modelling Document Usage in Competitive Intelligence Process. *International Conference on Transition from Observation to Knowledge to Intelligence (TOKI).* August 20 – 22, 2014. University of Lagos. Nigeria.

[2] Akanbi L. A. (2014) *Development of a Computational System for Integrating Usage into Document Indexing.* Ph.D. Thesis. Obafemi Awolowo University, Ile-Ife. Nigeria.

[3] Akanbi L. A., Adagunodo, E. R. and David, A. (2015) Integrating Document Usage with Document Index in Competitive Intelligence Process. *International Journal of Computer Applications.* Vol. 132. No. 13. Pp 37 - 43.

[4] Baeza-Yates, R. and Ribeiro-Neto, B. (1999). Modern Information Retrieval ACM Press New York.

[5] Camelo, C., Trigo, M. R., Quoniam, L. and Cardoso, J. C. (2011) Competitive Intelligence and the Development of Corporate Universities. In *Competitive Intelligence and Decision Problems* (ed A. David), John Wiley & Sons, Inc., Hoboken, NJ, USA. Chapter 14. Pp 281-298.

[6] David A. (2009) Relevance Information in Economic Intelligence. In *Information Science* (ed F. Papy). ISTE Ltd London and John Willey & Son New York

[7] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. and Harshman R. (1990): Indexing by Latent Semantic Analysis. *Journal of American Society for Information Science*. Vol 41. No 6. Pp 391-407.

[8] Dishman, P. L. and Calof, J. L. (2008) Competitive intelligence: a multiphasic precedent to marketing strategy, *European Journal of Marketing,* Vol. 42, No 7/8. Pp 766 – 785.

[9] Fur, N. and Buckley, C. (1991). A probabilistic Learning Approach for Document Indexing. *ACM Transactions on Information System*, *9*, pp223-248.

[10] Gupta, D., Bhatia, K. K., and Sharma, A. K. (2009). A novel indexing technique for web documents using hierarchical clustering. *International Journal of Computer Science and Network Security, 9*(9), pp 168-175.

[11] Gupta, P. and Sharma, A. K. (2010): Context Based Indexing in Search Engines using Ontology. *International Journal of Computer Applications (0975-8887)*. Vol. 1. No. 14. Pp 49-52.

[12] Menendez A., Atanes, E., Alonso, J., Merino, C., Bourgogne, P., Geffroy, P. *et al.*, (2002). *Economic intelligence. A guide for beginners and practitioners*, CETISME Partnership. Retrieved from http://www.madrimasd.org/Queesmadrimasd/Socios_Europeos/descripcionproyectos/Documentos/CETISM-ETI-guide-english.pdf (Retrieved 20/07/2012).

[13] Odumuyiwa, V. (2011) Collaborative Information Seeking in the Competitive Intelligence Process. In *Competitive Intelligence and Decision Problems* (ed A. David), John Wiley & Sons, Inc., Hoboken, NJ, USA. Chapter 4. Pp 69-91.

[14] Okunoye, O. B., David, A., and Uwadia, C. (2010). Amtea: Tool for creating and exploring annotations in the context of economic intelligence (Competitive Intelligence). In *11th IEEE International Conference on Information Reuse and Integration (IRI 2010)*, Pp 249- 252, Las Vegas, United States.

[15] [15] SCIP, (2012) About SCIP, *Strategic and Competitive Intelligence Professional.* Available at http://www.scip.org/content.cm?itemnumber=2214&navItemNumber=492, (Retrieved 15/10/2012

[16] Trigo, M. R., Gouveia, L. B., Quoniam, L., and Riccio, E. L. (2007) Using competitive intelligence as a strategic tool in a Higher Education context, *8th European Conference on Knowledge Management (ECKM). Consorci Escola Industrial de Barcelona (CEIB), Barcelona, Spain. 6-70* September 2007.

[17] Wang, P. and Soergel, D. (1998). A Cognitive Model of Document Use during a Research Project. Study I. Document Selection. *Journal of the American Society for Information Science.* 49(2), pp 115-133.

[18] Zaidi-Chtourou, S. and Bouzidi, L. (2010). Information Usefulness in an Information System: Performance at the Strategic Level of the Organisation. *Communications of IBIMA*. Doi:10.5171/2010.613429. Available at http://www.ibimapublishing.com/journals/CIBIMA/2010/613429/613429.pdf (Accessed 10/08/2012). Barcelona, Spain. 6-7 September 2007.

Authors' Briefs

Lukman Adewale Akanbi is a Lecturer at the Department of Computer Science and Engineering of Obafemi Awolowo University, Ile-Ife., Nigeria. He holds a B.Sc. degree in Computer Engineering in 2004. He also holds MSc and Ph.D. degrees in Computer Science from the Obafemi Awolowo University Ile-Ife, Nigeria. He is a registered Computer Engineer and a member of Nigeria Society of Engineer. His research area of interest includes: Intelligent system design, Human language processing and usage modelling of document and information.

Babajide Samuel Afolabi is a Lecturer at the Department of Computer Science and Engineering of Obafemi Awolowo University, Ile-Ife., Nigeria. He holds a B.Sc. degree in Computer Science from the Obafemi Awolowo University Ile-Ife, Nigeria in 1994. He also holds Ph.D. degree in Computer Science and Information Systems from the Université Nancy 2, Nancy, France in 2007. He is a registered Computer Professional and a member of Nigeria Computer Society. His research area of interest includes: Information systems, User modelling and Software Development.

Amos David is a Professor at the Université de Lorraine, Nancy France. He is a registered Computer Professional and the Secretary of the Executive Board of International Society for Knowledge Organisation (ISKO). He is also the President of the France Chapter of the Society. He has supervised many M.Sc. and Ph.D. theses both in France and Nigeria. He is a visiting Professor to a number of Universities in Nigeria, South Africa, Ghana and Benin Republic. His research area of interest includes: Information System, Competitive Intelligence and Knowledge Capitalisation.

Emmanuel Rotimi Adagunodo is a Professor at the Department of Computer Science and Engineering of the Obafemi Awolowo University, Ile-Ife., Nigeria. He is a registered Computer Professional and a member of Nigeria Computer Society. He has supervised many M.Sc. and Ph.D. theses. He is presently the Director of Intellectual Property and Technology Transfer Office of the Obafemi Awolowo University, Ile-Ife, Nigeria. His research area of interest includes: Operating Systems, Communication System and Information Storage and Retrieval System.