

Model-Based Information Retrieval System

Maitanmi OlusolaStephen^{a*}, Adekunle Yinka^b, Adio Adesina Kamorudeen^c, Onwodi Gregory^d, Malasowe Bridget^e

^aComputer Science, Babcock University Ilisan Remo, Ogun State. Nigeria.

^bComputer Science, Babcock University Ilisan Remo, Ogun State. Nigeria.

^cBasic Science, Babcock University Ilisan Remo, Ogun State. Nigeria.

^dSchool of Science. & Tech National Open Uni. of Nigeria

^eComputer Science, College of Education. Agbor. Delta State, Nigeria

^aEmail: maitanmi@yahoo.com

^bEmail: adekunleya@gmail.com

^cEmail: adesinaadio1@gmail.com

^dEmail: gregonwodi@hotmail.com

^eEmail: malasowebright@yahoo.com

Abstract

Information retrieval System (IRS) is often defined as the location and delivery of documents to a user to satisfy their information needs. IR is the area of study concerned with searching for documents, for information within documents, and for metadata about documents, as well as that of searching structured storage, relational databases, and the worldwide web. There is an overlap in the usage of the terms data retrieval, document retrieval, information retrieval, and text retrieval. IR is interdisciplinary; based on computer science, mathematics, library science, information science, information architecture, cognitive psychology, linguistics, statistics and law. Automated information retrieval systems are used to reduce what has been called information overload. Many universities and public libraries use IR system to provide access to books, journals and other documents. Web search engines are the most visible IR applications. This article is the design and implementation of information retrieval system that deals with this aspect of the institution's administration by providing an easy to use computerized application to assist in retrieving information.

Keywords-Information, retrieval, storage, metadata

*corresponding author Tel: +2348068856195
E-mail address: maitanmi@yahoo.com

1. Introduction

Information retrieval systems of most universities especially the third world countries have been a major concern that needs review. Student information should be kept properly for good documentation and proper retrieval to keep the institution's standard high and respectable status on the basis of information gathered.

Information retrieval or search plays an important role in a wide range of information management and electronic commerce tasks like Google in the Internet. Search also plays a key role in matching people and objects of Interest on commerce sites such as Amazon and eBay. In corporate intranets, institutional and governmental information sites, and personal computers, search is also an important means for enabling people to find information they need [1].

IRS is often defined as the location and delivery of documents to a user to satisfy their information needs. This simple task turns out to be highly complex. In order to facilitate research and move towards real systems, various pragmatic assumptions and simplifications have traditionally been made in IR research. The use of Information Retrieval is motivated by an information need. This information need can either be explicitly or implicitly verbalized. In a real world setting, the person seeking information formulates such a question and poses it to an expert. The expert calls upon his internal representation of the knowledge space and external documents and formulates answers. From the answers received the user extracts relevant points and gives feedback to the expert [2].

An information retrieval system is an application that stores and manages information on documents, often textual documents but possibly multimedia. The system assists users in finding the information they need. It does not explicitly return information or answer questions. Instead, it informs on the existence and location of documents that might contain the desired information.

The indexing process may include the actual storage of the document in the system and the information about the actual location of the document. Some suggested documents will, hopefully, satisfy the user's information need. These documents are called relevant documents. A perfect retrieval system would retrieve only the relevant documents and not irrelevant documents.

Information retrieval deals with the storage and representation of knowledge and the retrieval of information relevant to a specific user problem. Information retrieval systems respond to queries which are typically composed of a few words taken from a natural language. Users do not search just for fun; they have a need for information. The process of representing their information need is often referred to as the query formulation process [3].

The query is compared to document representations which were extracted during the indexing phase. The most similar documents are presented to the users who can evaluate the relevance with respect to their information needs and problems.

Although automatic indexing is widely used today, many information providers and even Internet services still rely on human information work.

There are three basic processes information retrieval system has to support:

- The representation of the content of the documents
- The representation of user's information need,
- The comparison of the two representations.

The processes are visualized in Figure 1.

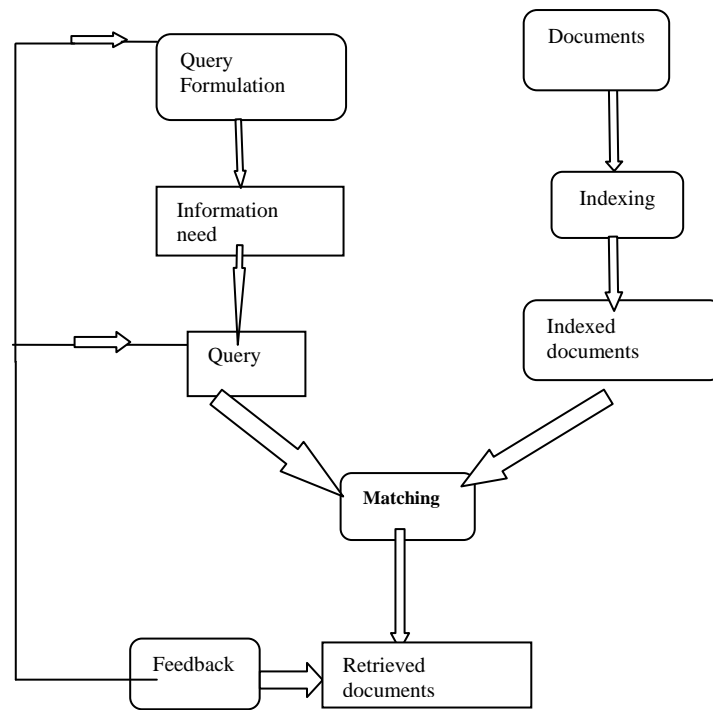


Figure 1- Information retrieval processes

Source: Research work, 2012

A. Objectives of the Study

The following points indicate the aims and objectives of IRS:

- i. The system shall improve the Information content of information resources collected.
- ii. The system shall enhance the Utility of information resources i.e. no redundancies of data.
- iii. The system shall improve the ability to identify, analyze and evaluate the information need of different groups and make informed decisions to satisfy them.
- iv. The System shall introduce the methods of obtaining feedback to users.
- v. The System shall improve the ability to create and use finding aids or retrieval tools.
- vi. The system shall introduce knowledge and understanding of the techniques and standards for Information Retrieval creation.

2. Literature Review

2.1. Historical Background of Problem Area

Research according to [5] emphasized that the idea of using computers to search for relevant pieces of information was popularized in the article 'As we may think' [6]. The first automated information retrieval systems were introduced in the 1950s and 1960s [6]. By 1970 several different techniques had been shown to perform well on small text corpora such as the Canfield collection (several thousand documents). Large-scale retrieval systems, such as the Lockheed Dialog system, came into use early in the 1970s. In 1992, the US Department of Defence along with the National Institute Standard and Technology (NIST), co-sponsored the Text Retrieval Conference (TREC) program. The aim of this was to look into the information retrieval community by supplying the infrastructure that was needed for evaluation of text retrieval methodologies on a very large text collection. This catalysed research on methods that scale to huge corpora. The introduction of web search engines has boosted the need for very large scale retrieval systems and even further.

2.2. Related Works

Over the years a lot of effort and approaches have been made to ease the search for information. Talking about information retrieval system, why the system is in existence initially is for the purpose of search and easy retrieval of information pertaining to the user's need. Various Systems to aid information search are Digital Library Archive tools such as E-

granary, Search engines; such as Google, Bing, Vertical search. Areas where information retrieval system was deployed will be examined below:

2.3. E-granary

The E-granary Digital Library provides those who lack adequate internet facilities an offline collection of approximately 30 million educational resources from 2000 websites and hundreds of CD-ROMS. The collection includes more than 60,000 books in totality, hundreds of full text journals and dozens of software application [7].

Through a process of copying Web sites (with permission) and delivering them to partner institutions in underserved areas of the world, this digital library delivers fast access to educational materials including video, audio, books, journals, and Web sites. These materials are delivered from Web servers connected to intranet Local Area Network (LANs) in the subscriber institutions.

Benefits of E-granary

1. It is cheaper, reduces organization's Internet cost.
2. High Speed of Processing.
3. It also provides training services on how to use the system.
4. It is easy to store and access information
5. Easily portable and usable where there is no Internet Service.
6. It is more reliable than Internet.

2.4. Google Search Engine

Google is now the most dominant search tool on web, setting the standards that others try to follow. Google was founded in the year 1988 relatively late compared to other search engines by Larry Page and Sergey Brin, who were graduates of Stanford university. Page and Brin had been working together on a search engine called "Backrub" since early 1996 but with the encouragement of Yahoo co-founder David Filo, they decided to start a company in 1998 and went looking for investors to back them.

Google was still in an alpha state, with an index or classification of just 25 million pages, but it was handling 10,000 search queries every day. The search engine and company grew quickly through word of mouth, initially with regular users coming across the tool and finding the results of their liking.

Benefits of Google Search Engine

Google is arguably the world's largest and most accurate search engine. What makes Google different from other search engines are:

1. It contains a Trillion Sites indexed:
2. It Provides Relevant Information:
3. It has Varieties of File Format:
4. Google Popularity:

2.5. Bing Search Engine

The first key is in the definition of a search engine. Literally speaking, we would define a search engine as a means to locate information contained within a defined information source [8].

Bing is a web search engine from Microsoft. Bing was unveiled by Microsoft CEO Steve Ballmer on May 28, 2009 at all things Digital Conference in San Diego for release on June 1.

Notable Changes include the listing of search suggestion while queries are entered and a list of related searches called explore pane based on the semantic technology from Power set which Microsoft purchased in 2008.

Benefits of Bing Search Engine

1. More relevant and accurate search results
2. Search history in view
3. Cleaner appearance makes browsing easier
4. Show just results for option
5. Brief simple descriptions

Limitations of Bing:

1. Fancy background images slow loading
2. Does not have scientific notation function of Google
3. Not different enough for consumers to warrant a change
4. Does not display broadly related topics

3. System Analysis and Design

The bulk of information in most universities resides in one form or another in a manual storage such as file folders in a file cabinets or desk drawers. Over the years, the number of records accumulated is, in some cases quite large, and consequently any retrieval of information is always tedious. With the current trend of software development, a vast number of systems have been developed using different software development methodologies, approaches, models, and techniques. One of such systems is the information Retrieval system of current article. This article is aimed at developing a system that will reduce information overload.

3.1 The Software Development Methodology

A development methodology is composed of a software development process model used in conjunction with one or more techniques.

This software was developed using Incremental lifecycle model. This methodology was adopted because it is most suitable for the software being developed.

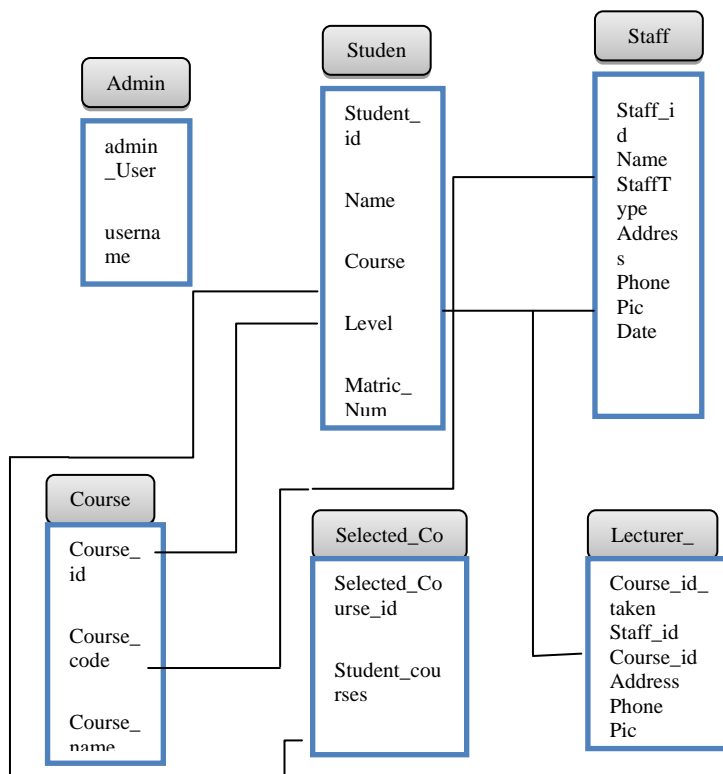


Figure 2- Schema representation from the proposed diagram: Source: Research work, 2013

3.2. Database

A database is an organized collection of data. The data is typically organized to model relevant aspect of reality. The term database system implies that data is managed to some level of quality and this in turn often implies the use of general purpose database management system [DBMS]. A general purpose DBMS is typically a complex software that meet many user requirements to properly maintain its database. Well known DBMS are Oracle, Sysbase, Microsoft Access, MSQl among others.

4. System Implementation and Review

This stage involves transforming theory into practices. The following two strategies are followed for running the system:

Parallel run: This is put into use for a certain defined periods; both the systems and manual are executed in parallel. This strategy is helpful because of the following:

Manual results can be compared with the results of the computerized system. Failure of the computerized system at the early stage, does not affect the working of the organization because the manual system continues to work, as it used to do.

Pilot run: In this type of run, the new system is installed in parts. Some part of the new system is installed first and executed successfully for considerable time period. When the results are found satisfactory then only other parts are implemented.

4.1. Programming platform

Based on the above specification, the choice of programming language for this article includes PHP, coupled with MySQL.

Hypertext Pre-processor (PHP). PHP is a server side programming language which is based on carrying out server requests and interaction with the database.

My Server Query Language (MySQL). MySQL is a free server query language compared to others in the likes of SQL, and it also supports a high traffic rate of users and allows simultaneous requests from multiple clients at the same time.

Meanwhile, it should be noted that all the necessary menus may not features here because of limited space but a highlight of the expected snapshot would be explained.

4.2. Deployment of the System

The system can be deployed using one of the methods explained above. The parallel switching method, the transition from the existing system to the proposed system is required after the system has been tested and proved to be efficient. Manual results can be compared with the results of the computerized system to detect the difference.

4.3. Unit and System Testing

This stage is the development cycle of this article, it seeks to put all the units that make this system into testing in order to make the system valid and also make sure that every part of it is working well as it is required of it.

4.4. User Interface Testing

This package helped in the design of the input and output units of the system. We tested the interfaces to ensure that they were properly linked to each other and that they interacted well to the database by retrieving the required data from the correct data fields provided in the interface.

4.5. The Operation of Information Retrieval Management System

In this section, we described the mode of operation of the developed system for universities. It shows the navigation through the pages and what each component contains as well as its functions.

4.6. Add new staff

This is the main menu form for staffs to input their details as required in the figure 3; this information is being stored in the database.

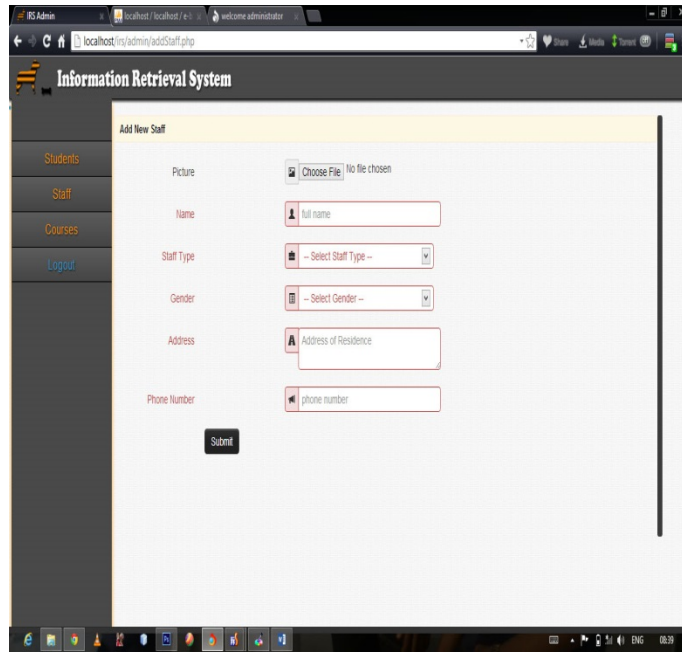
The image shows a web browser window displaying the 'Add New Staff' form of an 'Information Retrieval System'. The browser's address bar shows 'localhost:8080/admin/addStaff.php'. The system's header includes the title 'Information Retrieval System' and a navigation menu with options: 'Students', 'Staff', 'Courses', and 'Logout'. The main content area is titled 'Add New Staff' and contains the following form elements: a 'Picture' field with a 'Choose File' button and the text 'No file chosen'; a 'Name' field with a placeholder 'full name'; a 'Staff Type' dropdown menu with the text '- Select Staff Type -'; a 'Gender' dropdown menu with the text '- Select Gender -'; an 'Address' field with a placeholder 'Address of Residence'; and a 'Phone Number' field with a placeholder 'phone number'. A 'Submit' button is located below the form fields. The browser's taskbar at the bottom shows the system time as 08:39.

Figure 3. Adding New Staff

Source: Research Work, 2013

4.7. View all Courses

This is the main menu form where users get to view all the courses available as well as their course code and unit they carry as shown in figure 4 .

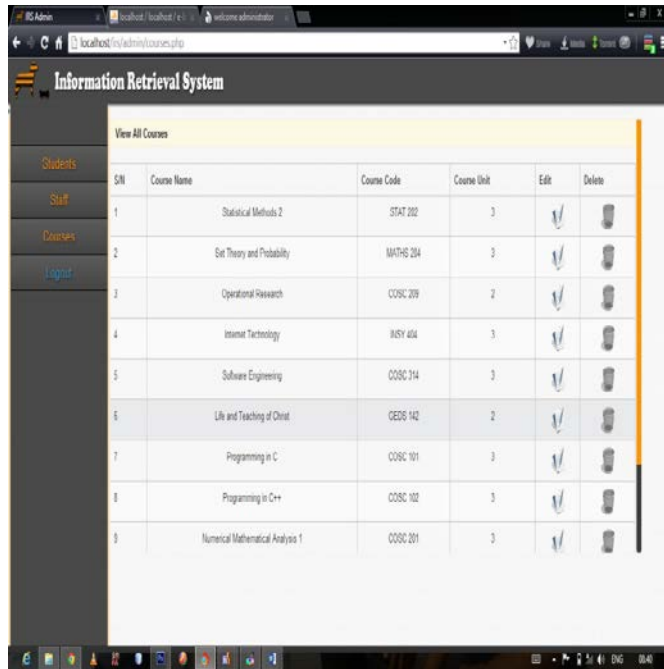


Figure 4 View selected courses.

Source: Research work, 2013

4.8. View all Students

This is the main menu form where users get to see all the students bio-data available from the database as well as their course of study, level, sex, age, matriculation number, hall of residence and their pictures as shown below in Figure 5



Figure 5 viewing all the students' records.

Source: Research work, 2013

Summary

A computer based information retrieval system involves the use of a computer to store, process and access both staff and student records. The application connects to a database where these records are kept and it has a user friendly interface that can allow a non-computer expert to use with ease. Construction of the system took place followed by the various testing carried out on the individuals units that make up the system. We tested the interface to ensure that they have maximum functionality. The database was also equally tested to ensure data integrity.

Conclusion

After a careful review of related works on information retrieval system, we were able to detect the limitations of the past works such as redundancy of data and improper management of memory space. We are able to design a small application which makes information retrieval easier. The research work has also exposed us to a lot of knowledge concerning the article, the activities involved in the system.

Recommendations

Regardless of the fact that the article has met the objectives of this research work, this research work still needs to undergo maintenance and improvement. The article is just an exercise to acquire knowledge; more features can still be added to the proposed system. Features such as speech retrieval to provide more semantics to information needs.

REFERENCES

- [1] S. Feldman NLP Meets the Jabberwocky (1999): "Natural language processing in Information Retrieval"; Online [Available]: <http://www.clips.ua.ac.be/clin2003/proc/03Brants.pdf> [Sept., 16, 2013].
- [2] S. R. Thomas, (1991) "Preliminary thoughts on an agent description language": *International Journal of Intelligent Systems*, Online: {available} [.www.onlinelibrary.wiley.com/journal/10.1002/\(ISSN\)1098-111X](http://www.onlinelibrary.wiley.com/journal/10.1002/(ISSN)1098-111X) [August, 10, 2013]
- [3] G. Tidhar, and J. Rosenschein, (1992) "A contract net with consultants" Online [Available]: [.www.cs.huji.ac.il/~jeff/papers.html](http://www.cs.huji.ac.il/~jeff/papers.html) [June, 12 2013]
- [4] S. Rao, M. P. Georgeff, and E. A. Sonenberg [1992] "Social plans": *Elsevier Science Publishers B.V.*: Amsterdam, The Netherlands, .A preliminary report. In E. Werner and Y. Demazeau, Decentralized AI 3 — Proceedings of the Third European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW-91) Online: [Available] <http://books.google.com.ng/books?isbn=0470519460> [July, 12, 2013]
- [6] S. R Thomas, (1995): "The PLACA agent programming language. In Intelligent Agents: Theories," Online [Available] <http://dl.citeseerx.ist.psu.edu/viewdoc/download?doi...> (eds M. Wooldridge and N. R. Jennings), LNAI Volume 890, pp. 355-369. Springer, Berlin. Architectures and Languages <http://dl.acm.org/citation.cfm?id=749460> [June, 10, 2013]
- [7] S. Rahman, M. Pipattanasomporn and Y. Teklu, Florida, (2007) "Intelligent Distributed Autonomous Power Systems (IDAPS)", *In Proc. the IEEE PES Annual General Meeting, Tampa*, Online: [Available] www.learningace.com, July, 01, 2013]
- [8] P. Agrawal, and Montreal, (2006) "Overview of DOE Microgrid Activities", Symposium on Microgrid,

Online: [Available] www.energy.ca.gov/.../conferences.../2006...23_microgrids_montrea... [August, 28, 2013]

[9] G. Weift,(ed.) (199) “Multi-Agent Systems”.*MIT Press, Cambridge, MA*.Online: [Available]

[www.ru.lv/.../Gerhard%20Weiss%20- %20Multiagent%20Systems%2..](http://www.ru.lv/.../Gerhard%20Weiss%20-%20Multiagent%20Systems%2..) [Sept., 24, 2013]

[10] R. Baez-Yates, and B. Rubiero-Neto, (1999) “Modern information retrieval, reading, MAAddison”*Wesley*.Online [Available]http://mail.im.tku.edu.tw/.../baeza-yates/chap10_user_interfaces [Aug. 12, 2013]

[11] C. J Van, and Rijsbergen (1979).“Information Retrieval. Butterworths” Online [Available]

www.onlinelibrary.wiley.com [June, 12, 2013]