Researchers Social Interaction over a Responsive Knowledge Based Network


Department of Computer Technology, Yaba College of Technology, Yaba, Lagos

Email address
kunledosu@gmail.com (Oladosu O. A.), vinctleye@gmail.com (Haastrup V. A.),
sade.mercy@yahoo.com (Okikiola F. M.), adeniranadebayo5585@yahoo.com (Adebayo A. A.),
saduwapat@yahoo.com (Ishola P. E.)

Citation

Abstract
Computing research relies heavily on the rapid dissemination of results. Hence, the formal process of submitting papers to journals has been augmented by other, more rapid, dissemination methods. Originally these involved printed documents, such as technical reports and conference papers. Then researchers started taking advantage of the Internet, putting papers on ftp sites and later on various web sites. But these resources were fragmented. There was no single repository to which researchers from the whole field of computing could submit reports, no single place to search for research results, and no guarantee that information would be archived at the end of a research project. This research paper gives attention to Research Portal (Research Repository or Research Management System) as a typical implementation of a responsive knowledge-based network.

1. Introduction

Information and communication technology (ICT) and its various applications have contributed immensely to the remarkable development of the society. A developed economy is as a result of full integration of ICT capabilities. It has led to the development of different web portals and knowledge based networks. The introduction of the portal concept to the Web has opened new possibilities to address some of the issues concerning the personal management of academic information and knowledge. Some of the main issues are the lack of integration, personalization and customization of existing academic information sources and systems.

A Web portal can be defined as a Web site that aggregates an array of contents and provides a variety of services including search engines, directories, news, e-mail and chat rooms. Portals have evolved to provide a customized gateway to Web information. A high level of personalization and customization is possible (Melzer 1999; Boye1999).

It was felt that the portal concept could be further developed to function as a sophisticated Web interface that can support the task performance (teaching and research) of academics.

Knowledge-Based Networking involves the forwarding of messages across a network based on some semantics of the data and associated Meta data of the message content. Knowledge Based Network tends to exploit the semantics of information when delivering information across a distributed environment only to those recipients who
have expressed a specific interest in that information. It has a model for the efficient filtering and dissemination of semantically enriched knowledge over a large loosely coupled network of distributed heterogeneous agents.

Using a semantically enhanced publish/subscribe model, based on content-based networking (CBN) removes the need to bind explicitly to all of the potential sources of that knowledge. Such a semantic-based CBN is called a Knowledge-Based Network (KBN), and I present this as the mechanism by which a distributed knowledge delivery service can be implemented. Examples of knowledge base networks are whatis.com, answer.com, wikipedia.org, google.com and so on.

A well-organized knowledge base can save enterprise money by decreasing the amount of time an employee spent trying to find information about - among countless possibilities such as tax laws or company policies and procedures and so on.

As a customer relationship management (CRM) tool, a knowledge base can give customers easy access to information that would otherwise require contact with an organization's staff and this makes the interaction simpler for both the customer and the organization. This is helpful in identifying sources, establishing notability, checking facts, and discussing what names to use for different things (including articles). Research Management System (RMS) simply called Research Repository or Research Portal, is a digital collection of the intellectual or research output. Institutional repository centralizes, collects, preserves and comply to open access concept of accessing collection of scholarly materials that showcase the research output of the institution or organization that make use of it. It incorporates a stored database of expert knowledge with couplings and linkages designed to facilitate information retrieval in response to specific queries, or to transfer expertise from one domain of knowledge to another. A knowledge base is a centralized repository for information; a public library, a database of related information about a particular subject. In relation to information technology (IT), a knowledge base is a machine-readable resource for the dissemination of information, generally online or with the capacity to be put online. As an integral component of knowledge management systems, a knowledge base is used to optimize information collection, organization, and retrieval for an organization, or for the general public. In general, a knowledge base is not a static collection of information, but a dynamic resource that may itself have the capacity to learn, as part of an artificial intelligence (AI), expert system. According to the World Wide Web Consortium (W3C), in the future the Internet may become a vast and complex global knowledge base known as the Semantic Web. The Semantic Web is an idea of World Wide Web inventor Tim Berners-Lee that the Web as a whole can be made more intelligent and perhaps even judicious about how to serve a user's needs. Berners-Lee observes that although search engines indexes much of the Web's content, they have little ability to select the pages that a user really wants or needs. He foresees a number of ways in which developers and authors, singly or in collaborations, can use self-descriptions and other techniques so that context-understanding programs can selectively find what users want.

2. Review of Earlier Research Works

In May 2000 a Web indexing workshop in the Netherlands decided to start an academic portal initiative in Europe. Main components will be: indexing and searching tools; cross-searching automatic indexes and human-made subject gateways; directory services; video-on-demand and streaming (Web Indexing Workshop 2000). The purpose of Campbell's white paper (Campbell 2000) on the scholar's portal is to suggest that the Association of Research Libraries (ARL) should seriously pursue the feasibility of developing a “library.org” Web presence. His paper refers to the proposed Web presence as the “scholars’ portal”. Increasingly the world’s business, including the business of research, is becoming Web based. Those agencies that wish to survive are busily developing new Web architectures and exploring how to migrate significant portions of their business to the Web environment. In the academic community, this move to the Web includes internal administrative business functions and increasingly the core functions of teaching and learning. Similarly, in the research library environment, integrated systems and digital library experiments have migrated to Web based functions almost totally. The scholar’s portal would promote the development of and provide access to the highest quality content on the Web. With the growing use of asynchronous learning methodologies, there is also an increasing need for extending certain elements of traditional library public services to the Web. This is already beginning to happen through experiments with virtual reference environments. A primary function of the scholar’s portal would be to provide researchers with an alternative means of retrieving dependable information beyond the capacity of commercial Web sites. Its goal would be to provide highly focused search engines adapted to the technical languages of the various academic specialties. By customizing search engines in this fashion and directing them to dependable sources of information, the portal would evolve increasingly “intelligent” automated systems and improve the success rate of query systems (Campbell 2000). An ARL Scholars Portal Working Group was set up in 2000 to explore how best to establish a collaborative research library presence on the Web. The Scholars Portal Project was launched in collaboration with a digital library software vendor in order to provide users with a single point of Web access that can reach a full array of diverse, high-quality information resources and deliver material directly to the user’s desktop (Quint 2002).

Research Management System (RMS) is a digital
collection of the intellectual or research output. Institutional 
respiratory centralizes, collects, preserves and comply to 
open access concept of accessing collection of scholarly 
materials that showcases the research output of the institution 
or organization that make use of it.

Research Portal is a knowledge-based portal designed for 
the Academic researchers. This allows you to add your 
research work to an array of many others research outputs 
after you must have registered. You can also search for 
research topics of your interest using Google search engine 
and you can search for topics already uploaded to the 
information system.

Research Portal- a broad-based open access, was founded 
on two key tenets: To publish the most exciting researches 
with respect to the subjects of our functional Journals. 
Secondly, to provide a rapid turn-around time possible for 
reviewing and publishing, and to disseminate the articles 
freely for teaching. There were many ideas and projects to 
support the task performance of academics before the arrival 
of the Internet and the World Wide Web.

In carrying out this research work relevant literatures in 
general research Portals and other specialized research 
management systems were reviewed. These include:

2.1. The Atlantic Online

The concept of hypertext and a memory extension came to 
life in July of 1945, when after enjoying the scientific 
companionship that was a side effect of WWII; Vannevar 
Bush’s “As We May Think” was published in The Atlantic 
Monthly. He urged scientists to work together to help build a 
body of knowledge for all mankind. Here are a few selected 
sentences and paragraphs that drive his point home.

Specialization becomes increasingly necessary for 
progress, and the effort to bridge between disciplines is 
correspondingly superficial or shallow.

The difficulty seems to be, not so much that we publish 
unduly in view of the extent and variety of present day 
interests, but rather that publication has been extended far 
beyond our present ability to make real use of the record. A 
record, if it is to be useful to science, must be continuously 
extended, it must be stored, and above all it must be consulted.

He not only was a firm believer in storing data, but he also 
believed that if the data source was to be useful to the human 
mind we should have it represent how the human’s mind 
works to the best of our abilities.

Our ineffectiveness in getting at the record is largely 
caused by the artificiality of the systems of indexing. ... 
Having found one item, moreover, one has to emerge from 
the system and re-enter on a new path.

The human mind does not work this way. It operates by 
association. ... Man cannot hope fully to duplicate this mental 
process artificially, but he certainly ought to be able to learn 
from it. In minor ways he may even improve, for his records 
have relative permanency. He then proposed the idea of a 
virtually limitless, fast, reliable, extensible, associative 
memory storage and retrieval system. He named this device a

memex. His description of the memex seems very familiar; 
consider a future device for individual use, which is a sort of 
mechanized private file and library. It needs a name, and to 
coin one at random, “memex” will do. A memex is a device 
in which an individual stores his books, records, and 
communications, and which is mechanized so that it may be 
consulted with exceeding speed and flexibility. It is an 
enlarged intimate supplement to his memory. (Bush 1945)

2.2. Gerard Salton (1960s - 1990s)

Gerard Salton, who died on the 28th of 1995, was the 
father of modern search technology. His teams at Harvard 
and Cornell developed the SMART informational retrieval 
system. Salton’s Magic Automatic Retriever of Text included 
important concepts like the vector space model (VSM) or 
term vector model (TVM) which is an algebraic model for 
representing text documents (and any objects, in general) as 
vectors of identifiers, such as, for example, index terms. It is 
used in information filtering, information retrieval, indexing 
and relevancy rankings.); Inverse Document Frequency 
(IDF): (The inverse document frequency is a measure of 
whether the term is common or rare across all documents); 
Term Frequency (TF): (The term frequency in a given 
document is simply the number of times a given term appears 
in that document); Term Discrimination Values: (Term 
Discrimination is a way to rank keywords in how useful they 
are for Information Retrieval.), and relevancy feedback 
mechanisms.

He authored a 56 page book called A Theory of Indexing 
which does a great job explaining many of his tests upon 
which search is still largely based.

2.3. Ted Nelson

Ted Nelson created Project Xanadu in 1960 and coined the 
term hypertext in 1963. His goal with Project Xanadu was to 
create a computer network with a simple user interface that 
solved many social problems like attribution.

While Ted was against complex markup code, broken 
links, and many other problems associated with traditional 
HTML on the WWW, much of the inspiration to create the 
WWW was drawn from Ted’s work.

2.4. Archie (1990)

The first few hundred web sites began in 1993 and most of 
them were at colleges, but long before most of them existed 
there was “Archie” which was the first search engine created 
in 1990 by Alan Emtage, a student at McGill University in 
Montreal. He created the program to search file names from 
the Internet. The original intent of the name was "archives,” 
but it was shortened to Archie. Archie helped solve this data 
scatter problem by combining a script-based data gatherer 
with a regular expression matcher for retrieving file names 
matching a user query. Essentially Archie became a database 
of web filenames which it would match with the users’ 
queries.
2.5. Veronica & Jughead

As word of mouth about Archie spread, it started to become word of computer and Archie had such popularity that the University of Nevada System Computing Services group developed Veronica. Veronica served the same purpose as Archie, but it worked on plain text files. Soon another user interface name Jughead appeared with the same purpose as Veronica, both of these were used for files sent via Gopher, which was created as an Archie alternative by Mark McCahill at the University of Minnesota in 1991.

2.6. File Transfer Protocol

Tim Berners-Lee existed at this point, however there was no World Wide Web. The main way people shared data back then was through File Transfer Protocol (FTP). If you had a file you wanted to share you would set up an FTP server. If someone was interested in retrieving the data they could use an FTP client. This process worked effectively in small groups, but the data became as much fragmented as it was collected.

2.7. Tim Berners-Lee & the WWW (1991)

While an independent contractor at CERN from June to December 1980, Berners-Lee proposed a project based on the concept of hypertext, to facilitate sharing and updating information among researchers. With help from Robert Cailiau he built a prototype system named enquire.

After leaving CERN in 1980 to work at John Poole's Image Computer Systems Ltd., he returned in 1984 as a fellow. In 1989, CERN was the largest Internet node in Europe, and Berners-Lee saw an opportunity to join hypertext with the Internet. In his words, "I just had to take the hypertext idea and connect it to the TCP and DNS ideas and — ta-da! — the World Wide Web". He used similar ideas to those underlying the Enquire system to create the World Wide Web, for which he designed and built the first web browser and editor (called WorldWideWeb and developed on NeXTSTEP) and the first Web server called httpd (short for HyperText Transfer Protocol daemon).

The first Web site built was at http://info.cern.ch/ and was first put online on August 6, 1991. It provided an explanation about what the World Wide Web was, how one could own a browser and how to set up a Web server. It was also the world's first Web directory, since Berners-Lee maintained a list of other Web sites apart from his own.

In 1994, Berners-Lee founded the World Wide Web Consortium (W3C) at the Massachusetts Institute of Technology.

Tim also created the Virtual Library, which is the oldest catalogue of the web. Tim also wrote a book about creating the web, titled Weaving the Web. Eventually, as it seemed that the Web might be profitable, investors started to get involved and search engines became big business.

2.8. Excite

This was introduced in 1993 by six Stanford University students. It used statistical analysis of word relationships to aid in the search process. Within a year, Excite was incorporated and went online in December 1995.

EINet Galaxy (Galaxy) was established in 1994 as part of the MCC Research Consortium at the University of Texas, in Austin. It was eventually purchased from the University and, after being transferred through several companies, is a separate corporation today. It was created as a directory, containing Gopher and telnet search features in addition to its Web search feature.

2.9. Jerry Yang and David Filo

Jerry Yang and David Filo created Yahoo in 1994. It started out as a listing of their favorite Web sites. What made it different was that each entry, in addition to the URL, also had a description of the page. Within a year the two received funding and Yahoo, the corporation, was Adm created.

Later in 1994, WebCrawler was introduced. It was the first full-text search engine on the Internet; the entire text of each page was indexed for the first time.

Lycos introduced relevance retrieval, prefix matching, and word proximity in 1994. It was a large search engine, indexing over 60 million documents in 1996; the largest of any search engine at the time. Like many of the other search engines, Lycos was created in a university atmosphere at Carnegie Mellon University by Dr. Michael Mauldin.

Infoseek went online in 1995. It didn't really bring anything new to the search engine scene. It is now owned by the Walt Disney Internet Group and the domain forwards to Go.com.

2.10. Alta Vista

Alta Vista also began in 1995. It was the first search engine to allow natural language queries and advanced searching techniques. It also provides a multimedia search for photos, music, and videos.

2.11. Inktomi Started in 1996 at UC Berkeley

In June of 1999 Inktomi introduced a directory search engine powered by "concept induction" technology. "Concept induction," according to the company, "takes the experience of human analysis and applies the same habits to a computerized analysis of links, usage, and other patterns to determine which sites are most popular and the most productive." Inktomi was purchased by Yahoo in 2003.
2.12. Sergey Brin and Larry Page

Google was launched in 1997 by Sergey Brin and Larry Page as part of a research project at Stanford University. It uses inbound links to rank sites. In 1998 MSN Search and the Open Directory were also started. The Open Directory, according to its Web site, "is the largest, most comprehensive human-edited directory of the Web. It is constructed and maintained by a vast, global community of volunteer editors." It seeks to become the "definitive catalog of the Web." The entire directory is maintained by human input.

At present, Google is a very popular search engine with a huge list of websites in their index. It is so popular that it leads to origin of the term "Googling," which means to search for information using Google. Bing is a search engine developed by Microsoft, which categorizes searches allowing for improved image and video searches along with preview searches.

This choice concerns the environment that the portal will be serving. The public portal, also called Internet portal, web portal or consumer portal, provides a single interface to the immense network of Internet servers. Its purpose is to attract the Internet community, because the larger the number of visitors the greater the probability of establishing virtual consumer groups that will potentially buy what portal advertisers want to sell. Since the middle 1990s, public portals have experienced three different stages of evolution: referential, personalized and interactive (Eckerson, 1999; cited in Dias, 2001), see also table 2. The public portal type is not relevant in this master thesis report, because the focus of this research lies on an internal HR portal within a case study organization and thus it is not a public portal accessible to other people who surf on the internet.

<table>
<thead>
<tr>
<th>Generation</th>
<th>Category</th>
<th>Level of public portal</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>Referential</td>
<td>Search engine, with hierarchical index of web content. Each index entry contains a description of the content object and a link to it. This generation emphasizes content management, mass dissemination of corporate information and decision support.</td>
</tr>
<tr>
<td>Second</td>
<td>Personalized</td>
<td>Through identification and a password, users create a Personalized view of portal contents, known as “Mypage”. This view shows just the categories each user is interested in viewing. The portal can notify users when new content is added to categories they have previously selected. Users can publish documents to the corporate repository so that other users may view them. This generation privileges content customized distribution.</td>
</tr>
<tr>
<td>Third</td>
<td>Interactive</td>
<td>The portal embeds applications that improve employees’ productivity, such as e-mail, workflow, project management, expense reports, calendars, schedules, etc. This generation adds collaborative character to corporate portals, providing multiple types of interactive services. Ports based on professional roles, for managing specific corporate functions, such as sales, human resources, finances, etc. This generation connects corporate applications with the portal, allowing users to execute transactions, read, write and update corporate data.</td>
</tr>
<tr>
<td>Fourth</td>
<td>Specialized</td>
<td></td>
</tr>
</tbody>
</table>

The corporate portal is an evolution from Intranets. This technology incorporates new tools that enable identification, capture, storage, retrieval and distribution of great amounts of information from multiple internal and external sources, useful for enterprise individuals and teams (Reynolds and Koulopoulos, 1999; cited in Dias, 2001). Corporate portals have also followed the same evolutionary stages experienced by public portals, though in a shorter period of time. These levels also influence the functionality and capability of the portal. It is believed that corporate portals have a potential to extend beyond the capabilities offered by public portals (Eckerson, 1999; cited in Dias, 2001) resulting in a fourth generation phase/category (specialized) (see table 2).

![Fig. 2. High-level architecture of a standard Web crawler](image-url)
3. Methodology

We attempt to examine different research repository to understand the mechanism and the general principles behind the design and functioning. Therefore, a research portal was developed to study the frequency of downloading, uploading, search for material and visit by researcher surfing through the website. We came up with various data which enables us to draw up a conclusion that a number of researchers are surfing the web to gain more knowledge and understanding into their topic of study and materials relating to their areas of research.

4. Research Repository and Search Engines

A search engine operates in the following order:
1. Web crawling
2. Indexing
3. Searching

Web search engines work by storing information about many web pages, which they retrieve from the HTML itself. These pages are retrieved by a Web crawler (sometimes called a spider) which is an automated Web browser that follows every link on the site. The contents of each page are then analyzed to determine how it should be indexed (for example, words are extracted from the titles, headings, or special fields called meta tags). Data about web pages are stored in an index database for use in later queries. A query can be a single word. The purpose of an index is to allow information to be found as quickly as possible. Some search engines, such as Google, store all or part of the source page (referred to as a cache) as well as information about the web pages, whereas others, such as AltaVista, store every word of every page they find. This cached page always holds the actual search text since it is the one that was actually indexed, so it can be very useful when the content of the current page has been updated and the search terms are no longer in it. This incorporates Google search for searching the World Wide Web to increase usability and satisfy user expectations that the search terms will be on the returned webpage. This satisfies the principle of least astonishment since the user normally expects the search terms to be on the returned pages. Increased search relevance makes these cached pages very useful, even beyond the fact that they may contain data that may no longer be available elsewhere.

When a user enters a query into a search engine (typically by using keywords), the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document's title and sometimes parts of the text. The index is built from the information stored with the data and the method by which the information is indexed. The engine looks for the words or phrases exactly as entered. Some search engines provide an advanced feature called proximity search which allows users to define the distance between keywords. There is also concept-based searching where the research involves using statistical analysis on pages containing the words or phrases you search for. As well, natural language queries allow the user to type a question in the same form one would ask it to a human. A site like this would be ask.com.

The usefulness of a search engine depends on the relevance of the result set it gives back. While there may be millions of web pages that include a particular word or phrase, some pages may be more relevant, popular, or authoritative than others. Most search engines employ methods to rank the results to provide the "best" results first. How a search engine decides which pages are the best matches, and what order the results should be shown in, varies widely from one engine to another. The methods also change over time as Internet usage changes and new techniques evolve. There are two main types of search engine that have evolved: one is a system of predefined and hierarchically ordered keywords that humans have programmed extensively. The other is a system that generates an "inverted index" by analyzing texts it locates. This second form relies much more heavily on the computer itself to do the bulk of the work.

Types of Research Portal
There are three major types of Research Portals namely
(i) Internet Research Portal
(ii) Intranet Research Portal
(iii) Local (stand-alone) Research Portal
(i) Internet Research Portal:
These are special sites on the web that are designed to help people find information stored on other sites. Internet Research Portals differ from other Research Portals in that they run on web servers or computers connected to the internet. Internet Research Portals are the powerful and sophisticated Research Portals with search capabilities reaching up to billions of web pages. They are fast, easy to use and more resourceful.


(ii) Intranet Research Portal
An intranet is a private computer network. It is the interconnection of network within an organization for the closed users of the organization. Most organizations today possess Intranet which enables security and safety to be ensured. In order to facilitate fast and easy location of information resources, intranet search engines are employed. Intranet research portal only search company intranet web pages and other documents when performing searches. Intranet Research Portals are cost-effective search solution for organizations using only information resources available within the company's Intranet. Examples are Web Glimpse, Harvest and Alta Vista Search Intranet.

(iii) Local (Stand-alone) Research Portal
These types of Research Portals are restricted in their search operations to information resources available within a stand-alone computer system. They are the least popular and most restricted type of search engines.

5. Result and Analysis

The graph below shows the number of post against the number responses in the table above. Number of Posts is at the Y-axis while Number of Response is at the X-axis. They are drawn in ascending order of Posts and Responses. The graph simply shows that there is tremendous high rate at which researchers surf the web and respond to topics like Education, Technology, Fashion, Electronics Engineering, Mobile devices, e.t.c. against topics in Automobile, Shares, e.t.c. based on professionalism, interest, applicability and usability of the available topics.

<table>
<thead>
<tr>
<th>Research Areas</th>
<th>Number of Posts</th>
<th>Number of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>15</td>
<td>30</td>
</tr>
<tr>
<td>Technology</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Fashion</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>Electronic Engg.</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>Automobile</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Shares</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Mobile Devices</td>
<td>18</td>
<td>33</td>
</tr>
</tbody>
</table>

Table 2. List of Posts and Responses during implementation process

Fig. 3. Pie Chart of the Research Areas and Number of Posts in 3-D

Fig. 4. Pie Chart of the Research Areas and Number of Response Posts in 3-D

Fig. 5. Multiple Bar Chart of the Research Areas, Posts and Number of Responses
6. Interpretation

All the charts and graphs actually show that there is a correlation between researchers, Topics Posted and Responses. Also there is a relationship between the number of response received and the topics posted i.e. the relative response time to any topic posted depends on the subject of discourse, how it affects the researchers and the researchers area of specialization. Hence, the rate at which users of research repository and researchers forum respond to topic posted is dependent on user’s interest and specialization with respect to the subject matter.

7. Summary and Conclusion

The study has been all about finding an easier, safer and educative ways of sharing opinions and interacts with other researchers in various part of the globe which can easily be referred to as Researchers Social Interaction (forum). In addition to chatting, this is a new and dynamics means of sharing ideas with different researcher all over the world within a flip of time. It has been discovered that Researcher Repository or Chat forum is one of the newest way in which Researcher and book authors of this world use now to reach out to their colleagues and numerous scholars in order to get feedback and responses concerning the entire areas of research topics to share of their idea and experience.

References


![Fig. 6. Graph of Posts against Responses](image-url)


