THE DETAILED NEED ASSESSMENT AND DEVELOPMENT OF FEDERATED SEARCH ENGINE APPLICATION ON ACADEMIC RESEARCH DATABASES OF ETHIOPIAN UNIVERSITIES TO CREATE RESEARCH INFORMATION SHARING ENVIRONMENT

BY

DULA BORU (B.Sc.)

A THESIS SUBMITTED TO THE DEPARTMENT OF INFORMATION SCIENCE, COLLEGE OF NATURAL SCIENCES, JIMMA UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTERS OF SCIENCE IN INFORMATION SCIENCE (ELECTRONIC AND DIGITAL RESOURCE MANAGEMENT)

OCTOBER, 2013

JIMMA, ETHIOPIA

JIMMA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCE

DEPARTMENT OF INFORMATION SCIENCE

ELECTRONIC AND DIGITAL RESOURCE MANAGEMENT GRADUATE PROGRAM

The Detailed Need Assessment and Development of Federated Search Engine Application on Academic Research Databases of Ethiopian Universities to Create Research Information Sharing Environment

Advisors

Michael B. Spring (PhD, Associate Professor)

Girum Ketema (MSc)

October, 2013

Jimma, Ethiopia

APPROVAL SHEET

This thesis entitled "The Detailed Need Assessment and Development of Federated Search Engine Application on Academic Research Databases of Ethiopian Universities to Create Research Information Sharing Environment" has been read and approved as the requirements of the Department of Information Science in partial fulfillment for the award of the Degree of Masters of science in Information Science (Electronic and Digital Resource Management), Jimma University, Jimma, Ethiopia.

Departmental Graduate Committee

Chairman	_____	_____
	Signature	Date
Michael B. Spring (PhD, Associate Professor), Advisor	_____	_____
	Signature	Date
Girum Ketema (M.Sc.), Advisor	_____	_____
	Signature	Date

ACKNOWLEDGEMENT

My first gratitude must go to my advisors, Michael B. Spring and Girum Ketema for all their guidance, advice, and valuable comments in developing this research project. Your dedication to timely commenting me is truly inspirational and I am honored to have had the opportunity to work with you.

I especially would like to thank Hunde Tekle and Dries for their helps and advice for the application development process.

I would never have been able to finish my thesis without the help from my friends, and support from my wife and family.

I would like to express my sincere gratitude towards Information Science department and Jimma University for giving me such chance of education.

ABSTRACT

Federated search (federated information retrieval or distributed information retrieval) is a technique for searching multiple text collections simultaneously. Queries are submitted to a subset of collections that are most likely to return relevant answers. The results returned by selected collections are integrated and merged into a single list. This research project aimed to identify the detailed need of federated search application and developing application of federated search tool on academic research in order to bridge databases in Ethiopian universities on research information sharing environment.

A descriptive cross sectional study design was used in this study. And data were gathered from three Ethiopian universities using questionnaire method of data collection. Stratified and purposive sampling techniques and Quantitative data analysis method was used for this study. The study finding confirmed that there is a high need of federated search engine application and also there is a problem of research information sharing between Ethiopian universities. Additionally the finding indicated that federated search engine application will widely use if it is developed and implemented on Ethiopian universities. Therefore in order to solve this need, this research project also delivers federated search engine application. Solr open source search engine software is used for the federated search engine application development.

KEY WORDS: Federated search engine, Solr, Indexing, Query, Result merging, Ethiopian universities.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVATIONS

AAU: Addis Ababa University

CGI: Common Gateway Interface

FIR: Federated Information Retrieval

FSE: Federated Search Engine

ICT: Information Communication Technology

IDF: Inverted Document Frequency

IR: Information Retrieval

JU: Jimma University

LAN: Local Area Network

SE: Search Engine

SSL: Semi-supervised Learning

WU: Wollega University

OPERATIONAL DEFINITIONS

Database - is an organized collection research data. The data is typically organized to model relevant aspects of reality, in a way that supports processes requiring this information

Federated search – a technique used for searching research collections simultaneously from multiple databases.

Indexing - it is the process of adding content to an index, this makes the contents searchable by the search engine.

Portal – a gateway or a point where users can start their search for information on the web.

Query –is a form of questioning, or in a line of inquiry.

Search engine – a tool used to crawl and index web pages, finding the best pages for specific lists of key words with good precision.

Solr- (pronounced "solar") is an open source enterprise search platform used to the search engine from the Apache Lucene project.

CHAPTER ONE INTRODUCTION

1.1 Background of the Study

In the electronic information environment one of the responses to the problem of bringing large amounts of information together has been for libraries to introduce portals. A portal is a gateway, or a point where users can start their search for information on the web. There are a number of different types of portals, for example universities have been introducing "institutional portals", which can be described as a layer which aggregates, integrates, personalizes and presents information, transactions and applications to the user according to their role and preferences. A second type of portal is a "subject portal"; a subject portal is an online gateway that helps users to quickly and efficiently find reliable, scholarly subject-based information in one place. A third type of portal is a "federated search tool" which brings together the resources to the library that subscribed to resources and allows cross-searching of these resources (Kumar, Sanaman & Rai, 2008).

Federated search (federated information retrieval or distributed information retrieval) is a technique for searching multiple text collections simultaneously. Queries are submitted to a subset of collections that are most likely to return relevant answers. The results returned by selected collections are integrated and merged into a single list. Federated search is preferred over centralized search alternatives in many environments. For example, commercial search engines such as Google cannot easily index uncrawlable hidden web collections while federated search systems can search the contents of hidden web collections without crawling (Kumar, Sanaman, & Rai, 2008).

In enterprise environments, where each organization maintains an independent search engine, federated search techniques can provide parallel search over multiple collections. There are three major challenges in federated search. For each query, a subset of collections that are most likely to return relevant documents are selected. This creates the collection selection problem. To be able to select suitable collections, federated search systems need to acquire some knowledge about the contents of each collection, creating the collection representation problem.

The results returned from the selected collections are merged before the final presentation to the user. This final step is the result merging problem.

One of the goals of this work is to assess the detailed need of federated search engine in Ethiopian universities to share their locally created content such as academic research output, lecture notes, guidelines, and reports. The need of federated search engine in Ethiopian universities was assessed in preliminary assessment done by the investigator and this work was used to assess the detailed need of the tool.

The other goal is developing federated search tool to facilitate sharing of these resources among universities if the need exists. The tools provide a single search interface to search contents of each universities local database and present a merged result for the user.

Currently, some of the universities in Ethiopia have their own research database that is used locally by their community and not shared with other institutions. So this research project is used as a bridge between universities in the country in order to share their research output with each other and it also helps students to refer the research works on all the databases. In addition to this, the tools can also be used by research advisors to find research topics which have already been worked by researchers in other universities. This way the tools contribute significantly in reducing plagiarism. The work is also used to initiate those universities who didn't start using online research databases to develop their own database and share with other universities in the county.

The thesis has two parts. First the research is undertaken in order to identify the detailed need of federated search engine by collecting data from selected universities. Different data collection mechanisms were used for the research and the data is analyzed to be used as an input for the project. Second, a federated search engine tool is developed based on the end result of the research.

1.2 Statement of the Problem

Information integration over distributed sources is an urgent problem to be solved for providing access to a variety of databases through a common search interface and portal (Kumar, Sanaman & Rai, 2008). There are many problems faced Ethiopian universities related to research information sharing.

As to my understanding, currently there is no literature or research work that indicates the need and the importance of research information sharing between Ethiopian universities and their impact on the quality of education and the research studies. The other problem is that there is no method or application that the students use to refer to researches conducted in other universities in order to use for their study. Also there is no way to disseminate researches in each university and their findings. This indicates a research conducted in one university is to be repeated in another university deliberately or not deliberately. Because, there is no way to identify either the research is already conducted or it is a new research work. Additionally, when the university research community wants to conduct a research on some problem, there is no application or method to check if the research is already conducted elsewhere. This results in the wastage of resources to conduct the research again.

Some of the universities in Ethiopia have their own local research database that is used to access research works and their findings within the university. But, they have to share this database with other universities and also access the research databases of those universities in order to develop their knowledge on every research conducted in universities in the country. As to my knowledge, there is no such interface that provides federated search engine to access those research works in every universities. There are 31 universities in Ethiopia and only few of them are using online research databases. So this research project is also used to initiates those universities who are not start to using online research database to use and gain the benefits of providing their research works to their students, staffs and others.

In summary, currently there is no mechanism to share research outputs of different universities which would help researchers in one university to get information about research outputs of other universities, prevent plagiarism among university students and avoid duplication of research works.

1.3 Objectives

1.3.1 General Objective

The general objectives of this research project is identifying the detailed need of federated search engine application in Ethiopian universities and developing application of federated search tools on academic research in order to bridge research databases in Ethiopian universities to create research information sharing environment.

1.3.2 Specific Objectives

- To assess how Ethiopian universities use research databases.
- To assess the detailed need of federated search application in Ethiopian universities in order to share their research databases with other universities.
- To develop federated search engine application based on the need assessment.
- To initiate those universities did not using online research database rather to develop their own online research repository.

1.4 Significance of the Research Project

In federated search systems, the task is to search a group of independent collections, and to effectively merge the results they return for queries. This work is used to assess the need and develop the tool that helps students and research communities to access research works in every Ethiopian universities using federated search engine.

The main significance of this research project is to help students by providing a reference and access to different research works in every university. This helps them to develop new research works based on the previous researches instead of repeating the research. This prevents the issue of plagiarism through checking the previously conducted researches on the system. The thesis also helps the instructors and advisors of the student research in the universities to identify either the student is working on a new research topic or the research was already conducted by other students in another university.

The other importance of this research project is to help a university research community by providing a single search interface to search every research works and their findings in all universities to use it as a reference for their study.

Additionally this thesis introduces a new system and it adds some knowledge and application on the current teaching learning process of Ethiopian higher educations by providing a new method of access to all researches in every university in the country. This helps the quality of researches and education given by the universities.

1.5 Scope of the Research Project

The scope of the research project is limited to the assessment of the detailed need of federated search tools in Ethiopian universities and developing an application of federated search engine based on open source search engine tool and tests on different databases. The assessment is done on selected universities of the country. There are 31 universities in Ethiopia and the population of the study is limited to three Ethiopian universities. They are Jimma University, Addis Ababa University and Wollega University. The reason to select those is that they represent three batches of Ethiopian universities. This study does not consider other universities due to time and financial resource limitations. Open sources software was used to develop the project. The project provides a single searching interface and it also includes indexing, metadata extraction, full-text searching and ranked results. The project is designed to run as a web application and it can be accessed on all major operating systems (Windows, Linux, and Mac and others).

1.6 Research Questions

The research project investigates the detailed need and user requirement for the tool that used to access researches in every university of Ethiopia. Based on the end result of the study the tool called federated search engine is developed in order to provide access to research databases of every university in Ethiopia to create research information sharing environment. Hence it tries to answer the need of federated search engine tool; the study is designed to answer the following questions:

1. Does the universities use online research databases?
2. How should research information available in local university database can be accessible from other universities.
3. Is federated search engine needed to be implemented in Ethiopian universities?
4. Does federated searching satisfy students' and the research communities' research information needs?
5. Does the federated search engine increase the quality of research conducted by students and universities research community?

1.7 Organization of the Thesis

This thesis contains six chapters. The first chapter gives information about the background to the search engine and federated search engine, statement of the problem, objectives, significance of the research project, and scope of the thesis and research questions of the research work. The second chapter deals with the literature review of federated search engine that are related to the objective of this thesis. The third chapter brings you to the details of methods and materials used for the need assessment study and the proposed methodology that were used to develop federated search engine application. The fourth chapter deals with the result findings and discussion of the research work. The fifth chapter discusses about the project implementation, features and processes of the project. In the six chapter of this paper conclusion and recommendation of the thesis is also covered. At last references of every documents used in this thesis is also listed.

CHAPTER TWO LITERATURE REVIEW

2.1 Overview of Federated Search Engine

One way or another all internet users use search engines to look for information on the web. A search engine is basically an information retrieval system designed to help finding information stored on a computer system or systems and the search results are usually presented in a list and are commonly called 'hits'.

Based on the general search engine, the new method of search engine called federated search is developed. Federated searching (sometimes known as broadcast searching, distributed searching, Meta searching, or parallel searching) is a technique used for searching collections simultaneously from multiple databases.

End-user federated searching of multiple databases stored by different companies in multiple locations is a relatively recent development. The majority of articles about today's federated search technology tend to fall into four categories. They (1) discuss the desirability and/or difficulty of creating a robust federated search tool (2) report on one or more specific federated search implementations, (3) compare federated search products currently on the market to each other and/or to Google Scholar, (4) look at how to implement a subject-specific federated searching Because these articles are theoretical, report of experience, or compare feature sets, they contain little data based on objective research.

2.2 Search Engine

Internet search is one of the most popular activities on the web. More than 80% of internet searchers use search engines for finding their information needs (Spink et al., 2006).

A search engine is basically an information retrieval system designed to help finding information stored on a computer system or systems (ICSTI, 2010). As such, search engines help reduce the time required to find information and also reduces the amount of information which must be consulted. It enables end users to target or focus on the few key relevant items.

It helps tackle the problem of ‘information overload’ which affects many areas of published information by bringing together, quickly, all relevant information in one succinct output or listing. To provide such a set of matching items, a search engine will typically collect metadata from a universe of items through a process of indexing. The index summarizes the main points about an item and requires a smaller amount of computer storage. Some search engines only store the indexed information and not the full content of each item, and instead provide a method of navigating to the item from the search engine result page. Alternatively, the search engine may and increasingly store a copy of each item as a full text item or a digital object.

Whereas some text search engines require users to enter two or three words in the search box separated by a space, other search engines may enable users to specify entire documents, pictures, sounds, and various forms of natural language. This is how search engines generally operate now. There has been a historical evolution which shows several distinct phases of development, and with each phase there has been a different set of players who have dominated the space.

In September 1999, Google claimed that it received 3.5 million queries per day. This number increased to 100 million in 2000, and has grown to hundreds of millions since the rapid increase in the number of users, web documents and web queries shows the necessity of an advanced search system that can satisfy users' information needs both effectively and efficiently (Shokouhi, 2007).

Since Aliweb (Koster, 1994) was released as the first internet search engine in 1994, searching methods have been an active area of research, and search technology has attracted significant attention from industrial and commercial organizations. Of course, the domain for search is not limited to the internet activities. A person may utilize search systems to find an email in a mail box, to look for an image on a local machine, or to find a text document on a local area network.

Commercial search engines use programs called crawlers (or spiders) to download web documents. Any document overlooked by crawlers may affect the user’s perception of what information is available on the web. Unfortunately, search engines cannot easily crawl documents located in what is generally known as the hidden web (or deep web) (Raghavan and Garcia-Molina, 2001).

There are several factors that make documents uncrawlable. Those factors are many hidden information sources only allow the access of their contents via the source-specific search interfaces due to intellectual property protection; and some information sources allow their contents to be copied by conventional search engines, but the information is updated very frequently and it is difficult for conventional search engines to crawl the updated information immediately; and the other factor is the access of the contents within some hidden information sources is subject to fee or subscription. A previous study (Bergman, 2001) has shown that the third type of information sources that require fee or subscription accounts for only about three percent of the whole hidden Web.

The main general purpose search engines crawl the Web and index Web pages, finding the best pages for each specific list of keywords with good precision. However, the so called deep Web contains information that is largely more valuable than the one that a current general–purpose search engine can discover.

2.3 Federated Search Engine

The development of new searching paradigms able to address more complex searches than those addressed to the current search engines and to discover deeper information is currently one of the most interesting challenges in the search computing field. Currently, the emerging paradigm is based on the combination of a multi–domain query approach with the integration of heterogeneous data sources capable to scour the deep Web. This has resulted in a new generation of search paradigms, called federated search engines (FSEs) that integrate search results from heterogeneous domain–specific content service providers (SeCo, 2006).

Federated information retrieval is a technique for searching multiple text collections simultaneously. Queries are submitted to a subset of collections that are most likely to return relevant answers. The results returned by selected collections are integrated and merged into a single list. Federated search is preferred over centralized search alternatives in many environments.

For example, commercial search engines such as Google cannot index uncrawlable hidden web collections; federated information retrieval systems can search the contents of hidden web collections without crawling.

In enterprise environments, where each organization maintains an independent search engine, federated search techniques can provide parallel search over multiple collections.

2.4 Need and Purpose of Federated Search Tools

The need and purpose of the federated searching are as follows: the growth of different types of databases, produced by different suppliers, with numerous interfaces and logins means that library users can find it confusing when attempting to access information (Kumar, Sanaman & Rai, 2008).

Library OPACs and web-pages have been alienating users with their use of library terminology and by including long lists of databases that users find it difficult to select from and search; the needs and expectations of library users, particularly students using academic libraries. The growth of different types of databases, produced by different suppliers, with numerous interfaces and logins means that library users can find it confusing when attempting to access information.

There are certain purposes which can be served by the federated search are as follows: transforming a query and broadcasting it to a group of disparate databases with the appropriate syntax. Merging the results collected from the databases, presenting them in a succinct and unified format with minimal duplication, providing a means, performed either automatically or by the portal user, to sort the merged result set.

In traditional search engines such as Google, only sources that have been indexed by the search engine's crawler technology can be searched, retrieved and accessed. The large volume of documents housed in databases is not open to traditional Internet search engines because of limitations in crawler technology. Federated searching resolves this issue by the technique described above and makes these deep Web documents searchable without having to visit each database individually (Kumar, Sanaman & Rai, 2008).

2.5 Search Models

Search on data bases starts with simple search formulation using single key term and combinations of terms. The search results thus obtained can be listed in simple predefined format or using user customized format to get desired information.

Various search models indicate the interactions of users search formulations and its interface with databases in a Library or online web access to aggregated databases (Kumar, Sanaman & Rai, 2008).

2.5.1 Old Search Model

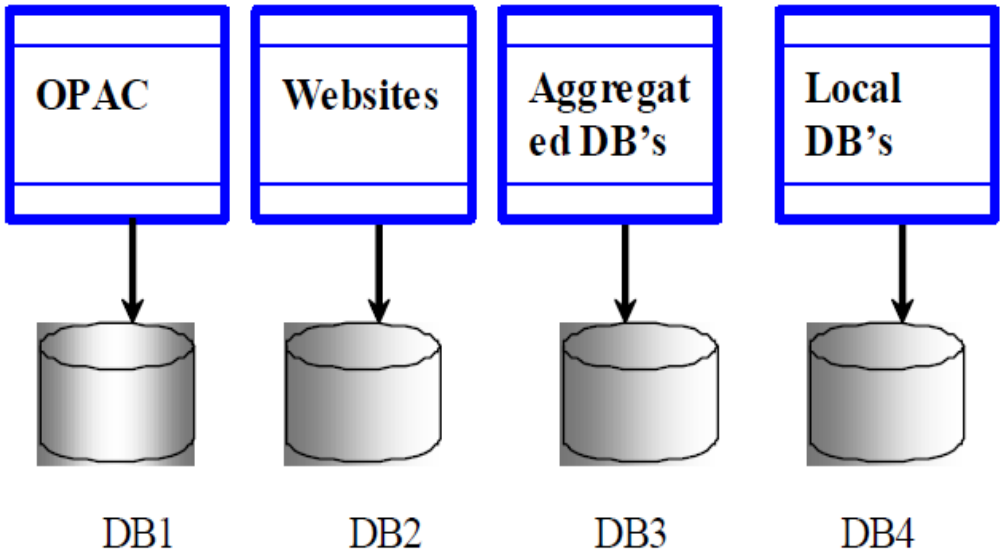

Figure 2-1 Old Search Model

Source: Adapted from (Kumar, Sanaman & Rai, 2008).

Without a federated search tool, each database requires its own unique search tool significantly complicating and slowing the search process for patrons. In this kind of search model, users submit their search query to individual database or search engines etc. and individually they get their required information. This kind of search tools are very time consuming and most of the time user didn't get their desired information.

2.5.2 Federated Search Model

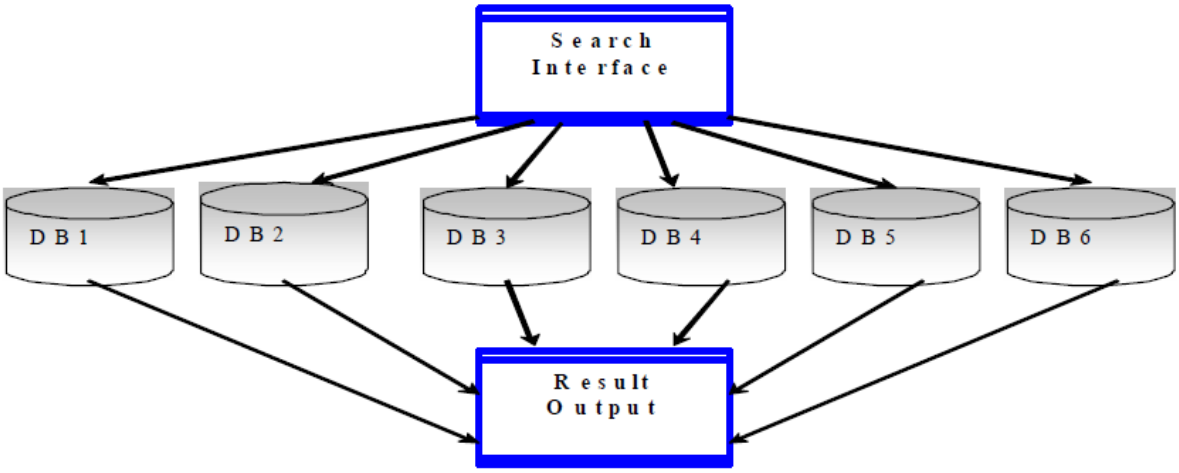

Figure 2-2 Federated Search Model

Source: Adapted from (Kumar, Sanaman & Rai, 2008).

Using federated search, a single tool searches and accesses all databases with one, easy to use interface. It provides opportunity to the users that they get their desired information from various databases, search engines etc.

2.6 How Federated Search Works?

Federated search computer programs allow users to search multiple data sources with a single query from a single user interface (Kumar, Sanaman & Rai, 2008). The user enters a search query in the portal interface's search box and the query is sent to every individual database in the portal or federated search list. Access details for the individual databases must be preset in the portal by its owner. Federated search systems either rely upon vendors to create commercial portal systems, or they rely upon government or other organizations to provide open access portals. How federated search is implemented depends upon which of the two types of organizations is providing the portal. Federated search portals, either commercial or open access, generally search public access bibliographic databases; public access Web-based library catalogues (OPACs), Web-based search engines like Google and/or open-access, government-operated or corporate data collections.

These individual data sources send back to the portal's interface a list of results from the search query. The user can review this hit list. Some portals will merely screen scrape the actual database results and not directly allow a user to enter the data source's application. More sophisticated ones will de-dupe the results list by merging and removing duplicates. There are additional features available in many portals, but the basic idea is the same: to improve the accuracy and relevance of individual searches as well as reduce the amount of time required to search for resources.

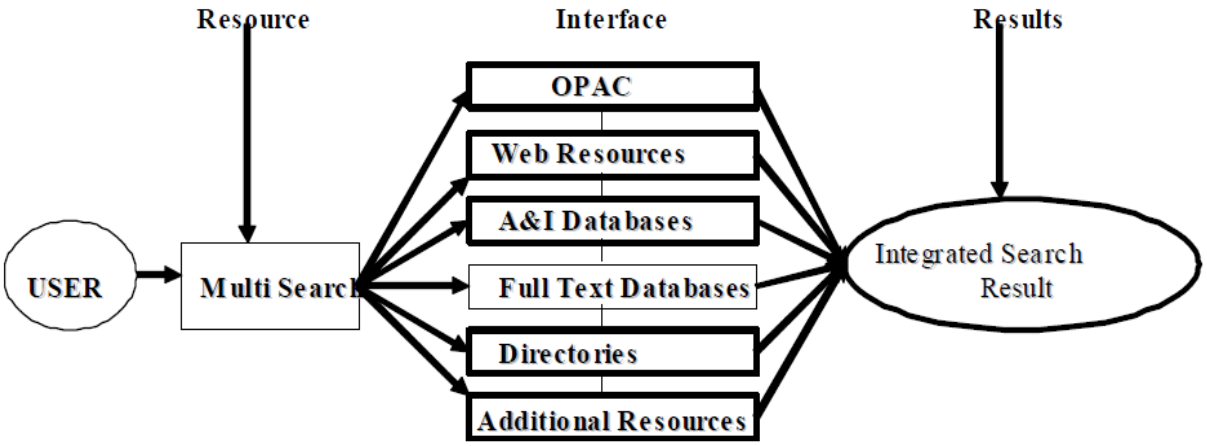

Figure 2-3 Diagrammatic representation of federated search engine

Source: Adapted from (Kumar, Sanaman & Rai, 2008)

In federated searching, a wealth of information is incorporated into single repositories that can be searched. In this model, the information is processed prior to the user's search. From the end-user's point of view, federated searching and Meta searching may seem similar, because both provide a single interface to multiple resources, but they actually differ in many respects. The pre-processing taking place in a federated searching environment, which we can describe as just-in-case processing, offers new opportunities regarding search methodologies and the presentation of results. For example, a ranking algorithm can be applied to each data element stored in the repository, unrelated to any future user query. Such an algorithm can take into account the number of times that an article has been cited, the number of articles that the author has published, the number of times that a book has been borrowed, a journal's impact factor, and other parameters.

A federated searching system can use the calculated rank to better evaluate the relevance of the specific item once it has been retrieved as the result of a query.

Federated searching: The system searches a local repository that was created earlier through the accumulation of data from numerous resources. Looking back a few years, we can see that the need for a single search interface to multiple resources arose some time ago, and, in fact, Meta searching and federated searching have been available for quite some time. Such systems originated in a variety of environments; for example, Elsevier, a publisher offering numerous journals, created a federated search mechanism enabling its user's to search all its e-journals through its Science Direct service. As Elsevier acquired other publishers, it was able to add their journals to the same platform.

Database vendors developed similar mechanisms. For example, Ovid provides a single interface to a few hundred databases that it publishes, and still retains them as separate databases. Commercial organizations were not the only ones that addressed the need for a single search interface; several large research institutions created a local environment based on federation. For example, the Los Alamos National Laboratory and the Ohio Link consortium in the United States, the University of Toronto in Canada, the Technical Knowledge Center of Denmark (DTV), and the Max Planck Society in Germany all offer large, diverse collections of e-journals that they store locally (Kumar, S., Sanaman, G. & Rai, N., 2008). These institutions have implemented federated searching to provide a single search interface across their electronic collections. However, not all organizations have the resources to adopt this just-in-case approach. Furthermore, with the rapid increase in the number of heterogeneous resources that institutions offer their users, a single federated searching system can serve only as a partial solution.

2.7 Federated Search Engine Implementation

When a federated search engine is implemented at a particular library, it then becomes a unique service. Federated-searching software allows customization, so no two implementations are exactly the same. For example, a library may choose to include all of its online resources as targets for a federated search engine or it may choose to create subject groupings first, each of which leads to a federated search service for a narrow topic.

Gerrity, Lyman, and Tallent discuss implementing a federated search system at Boston College, where they promoted the new service as "MetaQuest" (Kumar, Sanaman & Rai, 2008).

One application of federated searching is the metasearch engine; however, this is not a complete solution as many documents are not currently indexed. This is known as the deep Web or invisible Web. Many more information sources are not yet stored in electronic form. 'Google Scholar' is an example of a project trying to address this.

When the search vocabulary or data model of the search system is different from the data model of one or more of the foreign target systems the query must be translated into the each of the foreign target systems (Kumar, Sanaman & Rai, 2008). This can be done using simple data element translation or may require semantic translation.

2.8 Application of Federated Search

For visible Web contents, a previous study (Baeze-Yates & Ribeiro-Neto, 1999) has shown that users may prefer different search applications when they have different types of information needs. This is also true for federated search since there exist various federated search applications to satisfy different types of information needs, which may require different federated search applications.

The Complete Planet portal (Callan, Croft & Harding, 1992) provides structure guided browsing of thousands of hidden information sources. It enables users to explore a wide range of hidden information sources that they are interested in. This browsing model works well when users have broad information needs. However, when users' information needs can be easily expressed as text queries and when users want to directly find relevant information, other choices such as the information source recommendation application or the federated document retrieval application are more appropriate.

Information source recommendation (e.g., the Complete Planet portal and the IncyWincy invisible Web search engine) (Callan, Croft, & Broglio, 1995) goes a step further than the browsing approach by recommending most relevant information sources to information needs expressed as text queries.

This type of application is very useful if users want to browse the selected information sources by themselves instead of asking the system to retrieve relevant documents automatically. It is also a more appropriate choice when user interaction is required to choose from multiple search configurations for specific information sources.

An information source recommendation system is composed of two components, namely resource representation and resource selection. A more complex federated search solution is federated document retrieval. It selects relevant information sources for user queries, as does the information source recommendation system. Furthermore, user queries are forwarded to search the selected information sources and finally the returned individual ranked lists are merged into a single list to present to the users. Therefore, federated document retrieval provides a more complete search solution by combining all the three components of federated search: resource representation, resource selection and results merging. It is a more complicated solution than information source recommendation. Systems like Metalib (Callan, Lu & Croft, 1995) have been developed within cooperative environments, but very little has been pursued for uncooperative environments.

Some federated search applications include:

- Mednar.com - Searches medical information sources.
- Biznar.com - Searches business-related sources.
- WorldWideScience.org - Searches science content from all over the world, from government agencies, as well as other quality research and academic organizations.
- http://search.smartlib-bibliogen.ca/zengine?VDXaction=ZSearchSimple - Searches Capital Smart Library Consortium of Libraries.
- http://osulibrary.oregonstate.edu/metafind/about.html - Searches Oregon State University's Library.
- http://scienceroll.polymeta.com/search/ui7/searchfr.jsp?un=scienceroll - Searches a medical student's journey inside genetics and medicine through web 2.0.
- Science.gov - Searches science documents from a number of US federal government agencies.
- http://lifesearch.indexdata.dk/# - Searches University of Copenhagen's Library of Faculty of Life Sciences.

2.9 Benefits of Federated Search Engine

The major benefits of federated information retrieval (FIR) systems are to provide an effective search service over multiple collections. For a given query, the collections containing the most relevant answers are selected and then searched. The answers returned by all selected collections are then gathered and merged into a single coherent ranked list to present to the user.

The essential benefits of federated search to its users include efficiency, quality of search results, and current, relevant content (Leaderman, 2007).

2.9.1 Efficiency, Time Savings

Using a federated search engine can be a huge time saver for researchers. Instead of needing to search many sources, one at a time, the federated search engine performs the many searches on the user's behalf. While federated search engines specialize in finding content that requires form submissions to retrieve, it isn't the only criterion for being a federated search engine.

A federated search engine also associates content from different sources. Federated search uses just one search form to cover numerous sources, and combines the results into a single results page.

2.9.2 Quality of Results

Federated search engines show their value best in environments in which the quality of results matters, such as libraries, corporate research environments, and the federal government. In the case of the federal government, the constituents of the government benefit greatly from such applications.

A major difference between a federated search engine and a standard search engine like Google is that the client who contracts for the federated search service selects the sources to search. In almost every case, the sources will be authoritative. Google, on the other hand, has very minimal criteria for source selection.

If a Web page doesn't look like outright junk (i.e. Spam) Google will present it among the search results. Thus, the federated search engine acts as a helpful librarian does, directing users to excellent quality.

2.9.3 Most Current Content

In addition to filling out forms and combining documents from multiple sources, another important benefit of federated search engines is that they search content in real time. Real time data is crucial for researchers who are searching for up-to-the-minute content or for content those changes frequently. As soon as the content owner updates their source, the information is available to the searcher on the very next query. By contrast, with standard search engines/Google, the results are only as current as the last time that Google crawled sites with content that matches your search words. Content you find via Google might be days or weeks old, which can be fine depending on your situation, but can be problematic if you want the most current information.

2.10 Access Issues with Federated Search

Verification, authentication and certification can be difficult for the federated search vendor. Since federated search engines don't hold the data locally i.e. the engines perform the search, and send the results back to the portal (Kumar, Sanaman & Rai, 2008). The federated search engine must be able to access multiple password protected databases behind the scenes, or IP validate all at the one time, and show users their results in one easy navigable interface. The challenge for federated search vendors is to provide only licensed users access to databases, as specified for each license agreement that is in place for the organization.

2.10.1 Authentication

Authentication sets federated search engines apart from other expensive and highly sophisticated search software such as Verity and Autonomy. The latter usually restricts searches too internally generated information, ignoring subscription databases that enterprises have bought in-house.
All the user needs with federated searching is ID, password or IP validation along with files to be searched, and the federated search engine do the rest (Wilson, 2004).

2.11 Drawbacks of Federated Search

Limitations of the current generation federated search engines. These include:

- The lack of a uniform authentication standard means that some databases are inaccessible to federated search engines.
- True, full, de-duplication is impossible because databases download results in small sets and metadata standards vary by resource.
- Relevancy ranking is limited by the quality of the metadata, which usually does not include abstracts or full-text information.
- Although federated search systems are fundamentally software, they must be implemented and managed as a service, which takes a great deal of resources.
- Federated search engines cannot improve on the native interface in terms of search accuracy and precision.
- Federated searching is not for power searching needs. Just as with searching Meta search engines, only basic Boolean commands can be used.

The federated search has some other issues as well. First, it cannot cover all online library resources. The goal of one-stop shopping cannot be achieved completely by any federated search. There are various reasons for this:

- Some databases do not work with any federated search at all, such as SciFinder Scholar. SciFinder Scholar does not use a web browser but rather requires its own internet client. Neither Metalib nor WebFeat can cover SciFinder Scholar.
- If databases require a login, they will not work with the federated search.
- Some databases work with one federated search product but do not works with the other. Metalib cannot search LexisNexis databases because LexisNexis does not allow Z39.50 or XML gateway access. WebFeat cannot search databases that do not have a search box on their front page because WebFeat counts on the search box on the native interface to search.
- Many libraries have databases on a pay-per-search basis, and libraries normally do not want them to be searched by a federated search for budgetary reasons.

- Some databases have a limited number of concurrent users, and if these databases are included in a federated search, the limited seat(s) is/are taken immediately whenever someone logs into the federated search, and no other users can use these databases, libraries normally do not want to include databases with a very limited number of concurrent users in the federated search.

It may not make sense to add to a federated search menu the very specialized databases that most general users would not be interested in, or the databases that require special software. One example is Inter-university Consortium for Political and Social Research (ICPSR) that requires statistics software such as SPSS to view data.

2.12 Challenges of Federated Search Engine

There are three major challenges in federated search. For each query, a subset of collections that are most likely to return relevant documents are selected. This creates the collection selection problem. To be able to select suitable collections federated information retrieval systems acquire some knowledge about the contents of each collection, creating the collection representation problem. The results returned from the selected collections are merged before the final presentation to the user. This final step is the result merging problem.

2.12.1 Resource Description:

For resource description, previous research mainly focused on how to acquire corpus statistics of hidden information sources such as the vocabulary or term infrequencies (Callan & Connell, 2001) (Gravano et al., 1997).

However, in either the task of the information source recommendation system (to recommend information sources that contain as many relevant documents as possible) or the federated document retrieval system, the ultimate unit a user evaluates is a document. Therefore, it is necessary to estimate the characteristics of individual documents among the hidden information sources and is necessary to know how many documents each hidden information source contains (Liu et al., 2001).

For example, the resource selection algorithms need to estimate the number of relevant documents each information source contains and thus the information source size estimates are very important to adjust (normalize) the information source selection scores (Si & Callan, 2003a). However, information source size estimation is a major unsolved problem until now.

Previous research (Liu et al., 2001) required huge amount of communication costs to estimate information source sizes especially for large information sources. In this literature, a much more efficient Sample-Resample algorithm is proposed for this problem. This method utilizes sampled documents from query-based sampling and calculates the information source size estimates by sending resample queries and scaling the sampled document size with the ratio of document frequencies of 7 these queries in the whole information source to the document frequencies in the sampled documents (Si & Callan, 2003a).

2.12.2 Resource Collection

For resource selection, most prior research followed the "big document" strategy, which treats the information sources as "big documents" and calculates the similarities between user queries and the "big documents" to make the selection decision (Yuwono & Lee, 1997) (Callan, 2000) (Craswell, 2000) (French et al., 1999) (Xu & Croft, 1999) (Si et al., 2002b). However, as it is pointed out above that the ultimate units should be documents; the "big document" approach loses the boundaries between individual documents by simply treating an information source as a large document (Si & Callan, 2003a). This problem is serious as empirically studies have shown that the "big document" resource selection algorithms do not normalize the lengths of information sources well.

They often have strong disfavor bias against either small hidden information sources or large hidden information sources and thus miss large amount of relevant documents in these information sources (Craswell, 2000) (Si & Callan, 2003a) (Si & Callan, 2003c). In contrast, a resource selection algorithm is proposed to explicitly estimate relevant document distribution across available information sources for information source recommendation application by making full use of the information source size estimates and the content representations from the resource description component (Si & Callan, 2003a). This approach is not only more theoretically solid but also provides better empirical results.

2.12.3 Result Merging

Results merging are the last step for a federated document retrieval system, which merges the individual ranked lists from the selected information sources into a single final ranked list. It is a difficult job especially in uncooperative environments as different hidden information sources may use different retrieval algorithms or have different corpus statistics.

Previous results merging algorithms either used heuristic formula to calculate final comparable scores or assumed each hidden information source to return query term frequencies of the retrieved documents for computing consistent scores across information sources (Callan et al., 1995b) (Voorhees et al., 1995) (Kirsch, 1997). However, these methods are not very effective or require cooperation that is not valid in uncooperative environments. A Semi-Supervised Learning (SSL) results merging algorithm is proposed instead. It applies a centralized retrieval algorithm on the sampled documents acquired by query-based sampling.

The sampled documents with both information source independent scores and information source specific scores (returned from selected information sources) are used as training data. Linear models are learned from the training data to transform information source specific scores to corresponding information source independent scores. Furthermore, the linear models are applied on all the returned documents to approximate the comparable information source independent scores, and thus the final result list can be obtained with these source independent scores.

When there is not enough training data in the sampled documents, a variant of the SSL algorithm downloads a minimum number of documents “on the fly” to create additional training data (Si & Callan, 2002a) (Si & Callan, 2003b). The SSL algorithm has been shown to produce rather accurate final.

The solutions of the three sub- problems of federated search task are highly influenced by different environmental characteristics. In a small local area network such as small company environments, the information providers may cooperate to provide corpus statistics or use the same type of search engines (Callan, 2000) (Gravano et al., 1997) (Si et al., 2002b).

On the other side, in a wide area network such as very large corporate environments or on the Web there are many types of search engines and it is difficult to assume that all the information providers can cooperate as they are required (Si & Callan, 2002a) (Si & Callan, 2003a). Even they are willing to cooperate in these environments, it may be hard to enforce a single solution for all the information providers.

For example, a word in the stop word list of one information source which exists in almost every document may be quite indicative and cannot be thrown away for another information source and vice versa. Furthermore, it is often hard to detect whether information sources provide the correct information as they are required.

CHAPTER THREE METHODS AND MATERIALS

The work is aimed at studying the need of access to research databases of Ethiopian universities and to develop the tool that can be used to provide online access to those research databases through federated search engine application. This results in an easy access of research works of every university using single search interface. In this section, the details of methods and materials used for the need assessment and the proposed methodology that were used to develop the federated search engine application is discussed.

3.1 Study Area and Period

The study is conducted in three universities in Ethiopia. The first study area was Jimma University (JU) which is a public higher educational institution established in December 1999 by the amalgamation of Jimma College of Agriculture (founded in 1952), and Jimma Institute of Health Sciences (established in 1983). The two campuses are located in Jimma city 352 km southwest of capital city Addis Ababa. Jimma University has around 40,000 undergraduate and postgraduate students and 1200 academic staffs.

The second university studied was Addis Ababa University (AAU) which is established in 1950 and located in capital city of Ethiopia Addis Ababa. It has around 51,000 undergraduate and post graduates students and 2168 academic staffs. The third study area was Wollega University (WU) which was established in 2007 and located in Nekemte town 325 km west of Addis Ababa. Wollega University has around 10,000 undergraduate and postgraduate students and 400 academic staffs.

Data was gathered from April to March (April 15 to March 10, 2013).

3.2 Study Design

A descriptive cross sectional study design was used for this study, so that the data was gathered from the study population and analyzed in order to determine the detailed need of federated search engine application.

3.3 Source Population

The source population used for this study was all students and staffs of 31 Ethiopian universities.

3.4 Study Population

The study population of this study was all postgraduate students and academic staffs of Addis Ababa University, Jimma University and Wollega University.

3.5 Sampling Method

3.5.1 Sampling Technique

In this study stratified sampling technique was used to stratify the three universities so that data were gathered from each university proportionally. The reason for using stratified sampling technique is in order to obtain a representative sample from each three university. Under this technique, the population is divided into three universities, which is, each university is more homogeneous than the total population, that enable as to get more precise estimate for each stratum. Then samples are selected from the universities by random sampling technique.

3.5.2 Sample Size Determination

The total sample was selected from the total population of 7868, of which JU, AAU, and WU comprise 2850, 4258, and 760 respectively using stratified random sampling, where n is the total sample size, N the total population, Z is the probability value for standard normal distribution, P is the proportion of the need of federated search engine, and d was margin of error. The value of $Z_{\alpha/2}$ was 1.96 using 95% level of confidence, since no literature on the proportion of the need of federated search engine, P was 0.5, and the absolute margin of error d we use to have sufficient sample was 0.06. Then the total sample size n was obtained using;

$$n = \frac{n_0}{(1-1/N)+\frac{n_0}{N}} \quad ; \quad \text{Where} n_0 = \frac{Z_{\alpha/2}{}^2 * P*Q}{d^2} \qquad \ldots\ldots\ldots\ldots \qquad (1)$$

Based on the above formula (1), the total sample size was 283. Next we have to calculate the sample size in each stratum by taking postgraduate and academic staffs of Jimma university as stratum 1, postgraduate and academic staffs of Addis Ababa university as stratum 2 and postgraduate and academic staffs of Wollega university as stratum 3, based on proportional allocation to size in a way that helps us large sample was selected from larger number of population stratum and small sample was selected from smaller number of population stratum.

The sample size allocation (proportional allocation for JU, AAU, and WU) for each university was determined using statistical and proportion formula as shown below:

$$n_1 = \frac{n * N1}{N}, \qquad n_1 = \frac{283 \times 2850}{7868} = 103 \text{ for JU}$$

$$n_2 = \frac{n * N2}{N} \qquad n_2 = \frac{283 \times 4258}{7868} = 153 \text{ for AAU}$$

$$n_3 = \frac{n * N3}{N} \qquad n_3 = \frac{283 \times 760}{7868} = 27 \text{ for WU}$$

The sample size selected for each stratum was (103, 153, and 27) in JU, AAU, and WU respectively.

3.6 Instrument of Data Collection

In this study quantitative data was gathered using questionnaire as the main instruments of data collection. Data was collected from the selected sample study participants using primary method of data collection. Accordingly self-administered questionnaire was prepared and distributed to the selected sample of individuals.

To get enough information on the problem, closed and open ended questions were included in the questionnaire. Since the study focuses on the need assessment of federated search engine application in research information sharing environment, self-administered questionnaire was appropriate to get information about the need from post graduate students and academic staffs. Additionally Secondary sources of data such as books and scholarly articles on federated search engines were also used in the process of developing the project.

The validity and reliability of the instruments used for data collection was maintained before they are used to collect data by checking the questions in the instruments for their completeness, appropriateness and accuracy.

3.7 Variables

3.7.1 Independent Variables:

The independent variable of the study was:

- Usage of different university research databases
- Need of federated search engine.

3.7.2 Dependent Variables

The dependent variable of the study was:

- Age
- Gender
- occupation
- Computer usage
- Internet usage
- Usage of different search engines
- Usage of online research databases
- Application of federated search engine

3.8 Method of Data Analysis

The data collected using the above mentioned data collection instruments were cleaned by checking the filled questionnaires. After clearing the collected data, the quantitative data were entered in to a computer using SPSS software. Using SPSS, quantitative statistical data analysis were performed on the data. The statistical data analysis techniques were performed including descriptive statistical methods such as frequency table and charts to analyze the detailed need of federated search engine application.

3.9 Ethical Consideration

Ethical clearance was obtained from department of information science. Official letters were submitted to the three universities. All potential respondents were requested for oral or written consent prior to enrolment to the study. The purpose of the study was clearly described to the respondents including the benefits and risks of the study.

Participants' involvement in the study was on voluntary basis; participants who are unwilling to participate in the study and those who wish to quit their participation at any stage were informed to do so without any restriction. Any information concerning the study participant was kept confidential and the specimen collected from the study participants was only analyzed for the intended purposes.

3.10 Proposed Methodology of the Project

There are many articles, both academic and journalistic, proposing a methodology for Web development and federated search engine. Closer examination, however, shows these to be little more than ideas for best practice in designing the "look and feel" of a Web-site and federated search engine.

3.10.1 Phase One: Analysis

Phase One is concerned with the development of a federated search engine strategy and an analysis of how this search engine may achieve this strategy. It is known that the main reason for software project failures is misunderstanding of the system requirements. Phase one aims to reduce these risks by setting in place some strategic goals and objectives, and then designing a system to achieve them.

The decision to develop a federated search engine application should be based on the research result from the need assessment of federated search engine in order to create research information sharing between Ethiopian universities and the specifications and the look application that is developed. The output of this phase is used for the design of the project.

3.10.2 Phase Two: Design

Once the analysis phase has been completed, the development process can move on to the design phase, which is driven by the objectives of this project and the specification determined in analysis phase.

The design of this project can be broken down into two main tasks:

- Information Design: this task consists of three subtasks such as:
    - Resource selection
    - Resource representation and
    - Result merging
- Graphic Design: whereby the "look and feel" of the federated search engine application is designed for its intended audience. Search interface, screen layout, colors, images and animations and others are all designed during this step.

The output of the design phase are detailed design document that describes the structure of the federated search engine, the data structures of any databases that require development, and the functions of any CGI scripts required.

This design document represents the blue print of the application that are developed and used to practically develop the federated search engine application in the next phase.

3.10.3 Phase Three: Generation

The Generation of the project is focused on the generation of the federated search engine application and it is driven by the design document.

Step One: Resource Selection All the resources for the development of the federated search engine application, such as hardware, software, programming language, communications links and other required resources, were selected during this step. So, in order to develop federated search engine application the following resource were used. Those resources are computer, solr search engine software, network, database software's, web development software and other resources based on the design document.

Step Two: Design Review: During step two, the design document from phase two was compared with the available resources from the previous step to ensure the design can be achieved with the resources selected. If incompatibilities are found, the design phase and resource selection are reviewed. This is an iterative process, and if problems arise, phase one can be re-visited.

Step Three: Code Generation and Installation: The coding step sees the generation of codes of all of the functions specified in the design document using the selected programming languages and software connected with the application and its installation onto relevant web servers selected for the application. The installation step is simply posting the federated search engine application onto the web server, but it could also involve more complex tasks.

Step Four: Testing: Testing is one of the most complex and difficult areas of any application project. It is even more complex than with a traditional IS. Since this federated search engine applications are developed for a wide group of users in different technological environments, the application is tested against as many of these environments and combinations of technologies as possible in order to maximize the potential audience and the usage of the application to meet its objectives and goals.

CHAPTER FOUR RESULTS AND DISCUSSIONS

From the 283 distributed questionnaires 228 were filled and returned from all the three selected universities. During data cleaning three questionnaires were excluded due to higher number of (more than 10) missing values in each questionnaire. Hence, a total of 225 questionnaires collected were used for analysis of which 105 were from Addis Ababa University, 95 from Jimma University and 25 from Wollega University. This holds 79.5 % return rate. In this study the participation of the study population is divided in to postgraduate students and academic staffs and it was considered to be vital because the questionnaire has components which inquire the need of federated search engine application for their research work. Therefore, since the data collected properly addressed postgraduate students and academic staffs, the return rate (79.5 %) is reasonable to proceed with analysis.

4.1 Socio – demographic Characteristics of the Study Participants

Some of the socio – demographic characteristics of the study participants are presented in *Table 4.1* below. As it can be seen from the table and looking on the participant category of the respondent, most of the participants are postgraduate students constituting 54.38% of the study population and the remaining 45.62% is an academic staffs. In terms of the gender distribution of the respondents, majorities (85.09%) of the respondents are male and the remaining 14.91 % are female. Most of the respondents' are in the range of 26-25 age group. That is 55.26% are between 26-35 age group, 31.14% are in the range of 19-25 age group and the remaining 13.59% are between 36-50 age group were participated in the study.

Table 4-1 Socio-demographic characteristics of the study participants

Variables	**Classification**	**Percentages**
Respondents category	Postgraduate students	54.38%
	Academic staffs	45.62%
Gender	Male	85.09%
	Female	14.91%
Age Group	26 – 35 years	55.26 %
	19-25 years	31.15 %
	36 – 50 years	13.59

4.2 Computer, Internet and Search Engine Usage Assessment

The first objective of this study was to assess the use of research databases in Ethiopian universities. For that reason it is needed to assess the study population's computer, Internet and search engine usage frequency. As it is stated in the operational definition of terms in this document, Search engine (SE) is a tool used to crawl and index web pages, finding the best pages for specific lists of key words with good precision. Computer, Internet and Search engine usage of the study participants is presented in *Table 4.2* below. As it can be seen from the table most of the respondents (50.87%) uses computer for more than 5 hours per day, 34.21% of the respondents uses 3-5 hours per day, 11.84% uses 1-2 hours per day and the remaining 3.07% uses computer for less than 1 hour per day. The table also shows the frequency of internet usage of the study participants. So, 32.89% of the respondents' uses internet from 3-5 hours per day, 32.02% uses for 1-2 hours per day, 20.17% uses internet for more than 5 hours per day and the rest 3.07% uses internet for less than 1 hour per day.

The study also assessed the participants' access to popular search engines like Google, Yahoo and others. Accordingly, 35.96% of the respondents accesses search engines for 1-2 hours per day, 28.95% accesses for less than 1 hours per day, 25% accesses from 3-5 hours per day and the remaining 10.09% accesses those search engines for more than 5 hours in one day.

Table 4-2 Computer, Internet and Search engine usage of the study participants

Variables	**Classification**	**Percentages**
Computer usage of the respondents	More than 5 hours per day	50.87%
	3-5 hours per day	34.21%
	1-2 hours per day	11.84%
	Less than 1 hour per day	3.07%
Internet usage of the respondents	3-5 hours per day	32.89%
	1-2 hours per day	32.02%
	More than 5 hours per day	20.17%
	Less than 1 hour per day	14.91%
Search engine access of the respondents	1-2 hours per day	35.96%
	Less than 1 hour per day	28.95%
	3-5 hours per day	25%
	More than 5 hours per day	10.09%

4.3 Research Experience, Sources of Research Information and Frequently Used Search Engines

One of the main objectives of the study was to assess the use of Ethiopian universities research databases or repositories in order use them for their research work. So inquiring the participant's research experience and their sources of research information helps to assess the current problem of access to different research information sources.

The respondents' research experience, their source of research information and also frequently used search engines of the study participants are presented in *Table 4.3* below. Based on the response from the study participants, majorities of the respondents (87.72%) have a research experience and the rest 12.28% have no experience of research work. Additionally, the table below also lists the study participants' source of research information.

Accordingly, 74.56% of the respondent's uses web search engine such as: Google, Yahoo and others. And 13.16% of the respondents uses library as a source for research information, and 5.26% of the study participant accesses their university research databases to get useful information that helps them for their study and 4.38% accesses previous research collections at department/faculty and the remaining 2.63% uses other sources of information as their source of research information.

As it can be seen from the above result, most of the study participant accesses web search engines in order to refer or uses research information that used for research work. Therefore, the respondents frequently used web search engines are also presented in the table below. Based on the information in the table 93.86% of the study participants uses Google web search engine and 5.26% uses Yahoo search engine and the remaining 0.88% uses others search engines in order to find relevant information that is helpful for their study.

Table 4-3 Research experience, Source of Research information and frequently used search engines of the study participants

Variables	**Classification**	**Percentages**
Research experience of the respondents.	Yes	87.72%
	No	12.28%
Research Information sources of the respondents	Web search engines	74.56%
	Library	13.16%
	University Research databases	5.26%
	Department or Faculty	4.38%
	Others	2.63%
Frequently used web search engines of the respondents.	Google	93.86%
	Yahoo	5.26%
	Others	0.88%

Based on the results in the table above there is a problem of access to university research databases. The main reason for this problem is that only few universities in Ethiopia have their own research database or repository. Even in those universities that have research repositories, the access is very low when compared to popular web search engines. The reason behind this is that there is a lack of awareness to a user's. Additionally one of the main objectives of this study was to initiate those universities not using online research database to develop their own online database. So, those universities should have to develop their own research databases or repositories in order to provide access to the collection of students and staff researches in their university.

4.4 Research Advising Experience and the Agreement to the Student Research Plagiarism Practice

As it was mentioned above, the study participants of this study are divided in to postgraduate students and academic staffs of three selected Ethiopian universities. In order to assess the research advising experience of the respondents, only respondents from academic staff is used.

The research advising experience of the respondents and their agreement to the student copying research from other universities practice is presented in *Table 4.4* below. Additionally, either the respondents use any method or application to check or verify the copied research from a new research work or not is also listed in the table. Based on the results from the response of the study participants most of the respondents advise students researches. That is 98.1% of the study participant have research advise experience and the remaining 1.9% have no experience of research advising. As it was mentioned in the statement of the problem section of this document, currently there is no way to disseminate researches in each Ethiopian university and their findings. This results the research conducted in one university is repeated in another university, because some of the students are copying their research from other universities.

As it can be seen from the table below, most of the respondents agreed that the students copy their research from other universities either a full research content or partially. That is 91.35% of the study participants agreed that there is a problem of student research plagiarism issues in Ethiopian universities and the remaining 8.65% disagreed.

Even though, most of the respondents agreed on the student research plagiarism problem, they respond that there is no method or application to verify either the research is new work or it is copied from other universities. Based on the results in the table below 81.7% of the study participants responded that they didn't use any method to check or verify the originality of the student's research work. But the remaining 18.3% manually verify during the advising and evaluation of students research.

Table 4-4 Research advice experience and respondents agreement to the students copying their research from other universities practice of the study participants

Variables	**Classification**	**Percentages**
Research advises experience of the respondents.	Yes	98.1%
	No	1.9%
Respondent's agreement to the students copying their research from other universities practice.	Agree	91.35%
	Disagree	8.65%
Method or application to check either research is new work or copied from other Universities.	No	81.7%
	Yes	18.3%

The results in the table above shows that there is a problem of students research plagiarism or the duplication of researches in Ethiopian Universities. The main reason for this problem is that there is no application that provides online access or used to disseminate those research work and their findings in each university.

Additionally, as long as there is no method or application to verify, knowing where the students steal researches from is also a difficult task. This shows that there is a need of application that used for access to every universities research repositories.

4.5 Access Exposure and Agreement to the Necessity of Accessing Other Universities Research for the Respondents Study.

The exposure of the respondents' access to other universities researches works and their agreement to the necessity of those researches to their study is presented in *Table 4.5* below. According to the study participant's response on their access exposure to other universities research works, most of the respondents (93.5%) have no exposure and experience of access to those researches and the remaining 6.5% accesses other universities research through their own different mechanisms.

The study also assessed the necessity of other universities research works and findings to the respondents study. Majority of the study participants believe that accessing other universities researches are essential for their research in order to use them as a reference and starting point for their study. Based on the results 95.61% of the total study participants responded that other universities researches are necessary for their study and research work and 4.39% answered that it is not necessary for their research.

Table 4-5 Respondents access exposure and their agreement to the necessity of accessing other Universities researches for their study

Variables	**Classification**	**Percentages**
Respondents' exposure of Accessing other universities research works.	No	93.5%
	Yes	6.5%%
Respondents Agreement to the necessity of other universities researches for their study.	Yes	95.61%
	No	4.39%

Based on the results in the table above most of the respondents want to access other universities research and also those researches are necessary for their study to use them as a reference. But currently there is no application that provides this service to those users. One of the objectives of this study is to assess the need of federated search engine. The result above indicates that there is a need of an application that provides access to every university research repositories.

For this reason, this research project is needed to solve the need of application by developing federated search engine in order to give access to search across multiple research databases using a single search interface.

4.6 The Agreement to the Impact of Lack of Application to Access Other Universities Researches and Research Information Sharing Problem

In this study, in order to identify the need of federated search engine application, the respondent's agreement to the current impacts of the lack of application and the research information sharing problem was assessed, Because they shows the current gap and need of the main objective of this research project. The *Table 4.6* below lists the study participants' response to the above mentioned two problems. As it can be seen from the table 92.3% of the total respondents agreed that the lack of application has an impact on their research work and the remaining 7.7% responded that it has no impact on their study. In the other hand, the study also assessed the respondents' agreement to research information sharing problem between Ethiopian universities. Regarding to this problem 92.1% of the study participants agreed that there is a problem and the remaining 7.9% disagree about the problem of research information sharing between universities.

Table 4-6 The Respondents agreement to the impact of lack of application to access other Universities researches on their study and research information sharing problem between Ethiopian Universities

Variables	**Classification**	**Percentages**
Respondents agreement to the impacts of lack of application to access other universities researches on their research work	Agree	92.3%
	Disagree	7.7%%
Respondents Agreement to the research information sharing between Ethiopian universities.	Agree	92.1%
	Disagree	7.9%

The result in the above table shows that the lack of application has a strong impact on the study participants research works. Additionally, the respondents are also agreed that there is a research information sharing problem.

This indicates that federated search engine that used to provide online research information sharing between the universities should be developed and implemented on Ethiopian universities to solve the above mentioned problems.

4.7 The Need for Federated Search Engine and the Degree of Usage for the Application if it is implemented

Federated search is the process of performing a simultaneous real-time search of multiple diverse and distributed sources from single search page, with the federated search engine acting as intermediary. That means it is a method or an application used to retrieve or search from multiple data sources at the same time using a single search interface. The main objective of this research project was to assess the need of federated search engine in order to access different universities research databases and developing an application that provide this service.

The study participants are directly asked either there is a need for federated search engine application or not. As it can be seen in the *Table 4.7* below, most of the study participants (65.35%) responded that the need for this application is very likely, 25.88% need is likely and the remaining 6.14% need of federated search engine application is neutral.

The other thing presented in the table below is how the study participants uses if this application is implemented on Ethiopian universities in order to deliver access to other universities research information. Accordingly, 53.51% will use the application very extensively, 35.09% will use it extensively and the other 11.4% responded neutral in their use of this application.

Table 4-7 Respondents need for federated search engine and their degree of usage for the application if it is implemented on Ethiopian Universities

Variables	**Classification**	**Percentages**
Respondents need for federated search engine application	Very Likely	65.35%
	Likely	25.88%
	Neutral	6.14%
	Unlikely	2.63%
Respondent's degree of usage for federated search engine application if it implemented on Ethiopian universities.	Very Extensively	53.51%
	Extensively	35.09%
	Neutral	11.4%

The result in the above table shows that there is a high need of federated search engine application and the respondents will use it widely if it is developed and implemented. Therefore it confirms that the federated search engine application should be developed and to be implemented on Ethiopian universities to solve the above mentioned problems.

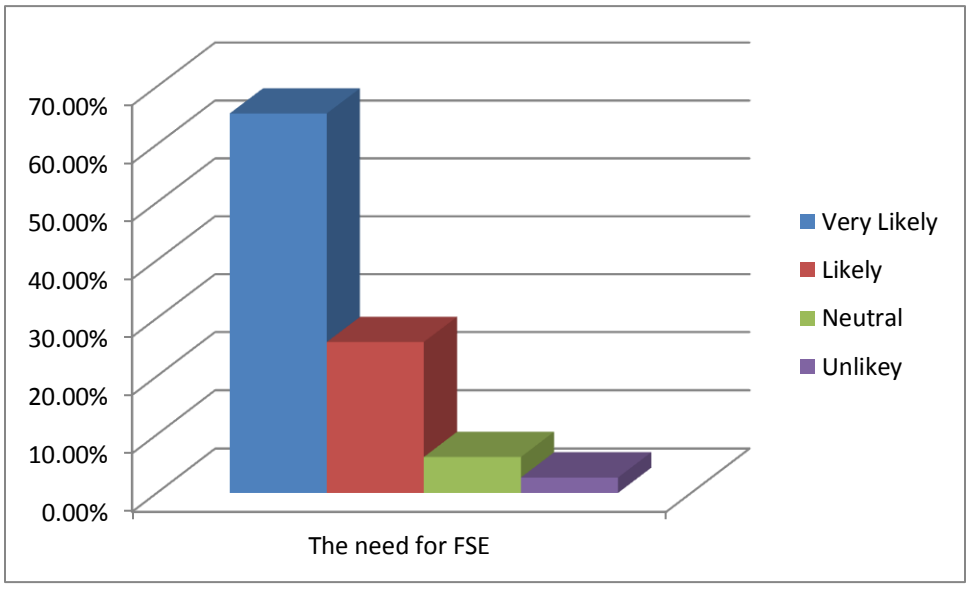

Figure 4-1 Respondents need for federated search engine application

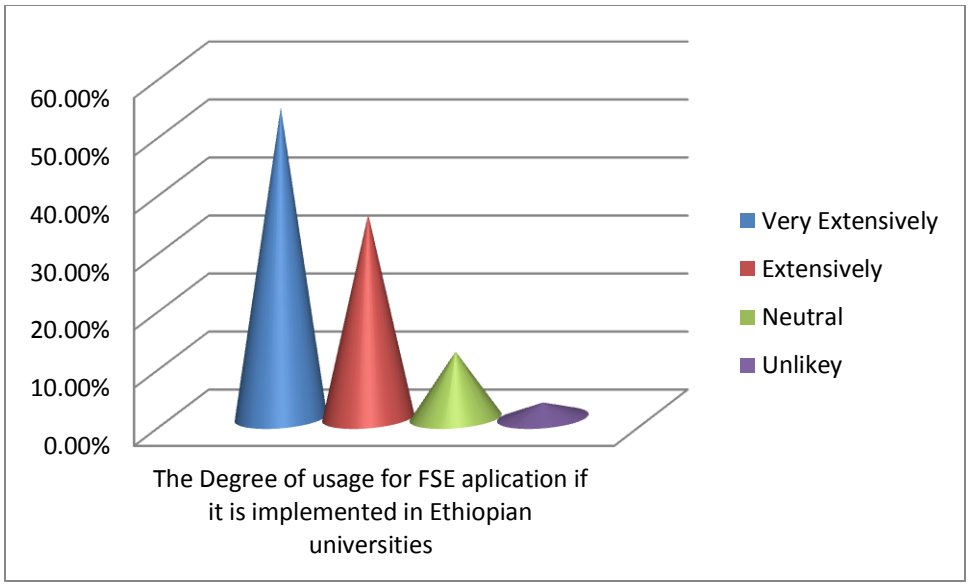

Figure 4-2 Respondents degree of usage for FSE application if it is implemented on Ethiopian universities

CHAPTER FIVE PROJECT IMPLEMENTATION

5.1 Overview of the project

With the amount of online information rapidly expanding and residing in increasingly disparate sources, organizations need a way to simplify how their users discover and access the information they need. Federated search is designed to help organizations meet this challenge, enabling users to simultaneously search multiple sources and quickly obtain relevant results using a single search query.

As it was mentioned in the objective part of this document, one of the objective of this research project was to develop federated search engine application that provide an online access to multiple research databases of Ethiopian universities based on the results of the study above. Accordingly, this project is developed based on the need assessment of federated search engine application in Ethiopian universities that used to share their research information. This project came about to meet the need of searching multiple content sources with one query. This allows a user to search multiple research databases at once in real time, arrange the results from the various databases into a useful form and then present the results to the user. The project is developed as a prototype by using Solr open source search engine software and it provides a single search interface to search three different databases at the same time and deliver a merged result to a user.

The benefits of the project is that it provide one-stop access to multiple research information sources and users don't need to know where or how to search, it saves researcher time and money and improves utilization of information sources. Additionally it provides consolidated, de-duplicated results and important results are not missed.

5.2 Solr Search engine application

Solr (pronounced "solar") is an open source enterprise search platform from the Apache Lucene project. Its major features include full-text search, hit highlighting, faceted search, dynamic clustering, database integration, and rich document (e.g., Word, PDF) handling, providing distributed search and index replication. Solr is highly scalable and it is the most popular enterprise search engine (Apache solr reference guide, 2013).

Solr is written in Java and runs as a standalone full-text search server within a servlet container such as Apache Tomcat or Jetty. Solr uses the Lucene Java search library at its core for full-text indexing and search, and has REST-like HTTP/XML and JSON APIs that make it usable from most popular programming languages. Solr is powerful external configuration allows it to be tailored to many types of application without Java coding, and it has a plugin architecture to support more advanced customization.

Solr is based on the Apache Lucene project, a high-performance, full-featured search engine. It offers support for the simplest keyword searching through to complex queries on multiple fields and faceted search results. Some of the Solr features are:

- Advanced full-text search capabilities
- Optimized for high volume web traffic
- Comprehensive html administration interfaces
- Flexible and adaptable with xml configuration
- Extensible plugin architecture
- Sort by any number of fields
- Faceted searching based on unique field values, explicit queries, or date ranges
- Multi-select faceting by tagging and selectively excluding filters
- More like this suggestions for given document
- Spelling suggestions for user queries
- Auto-suggest functionality
- Allow configuration of top results for a query, overriding normal scoring and sorting.
- Full html admin interface

A Solr index can accept data from many different sources, including XML files, comma-separated value (CSV) files, data extracted from tables in a database, and files in common file formats such as Microsoft Word or PDF.

5.3 Architecture of the Project

The *figure 5.1* below shows the architecture of the application. The architecture consists of four components. Those components are clients, Webserver and the federated search engine, that is a core of the applications that perform indexing and querying every document in the database. The other component of the system is the document store or databases.

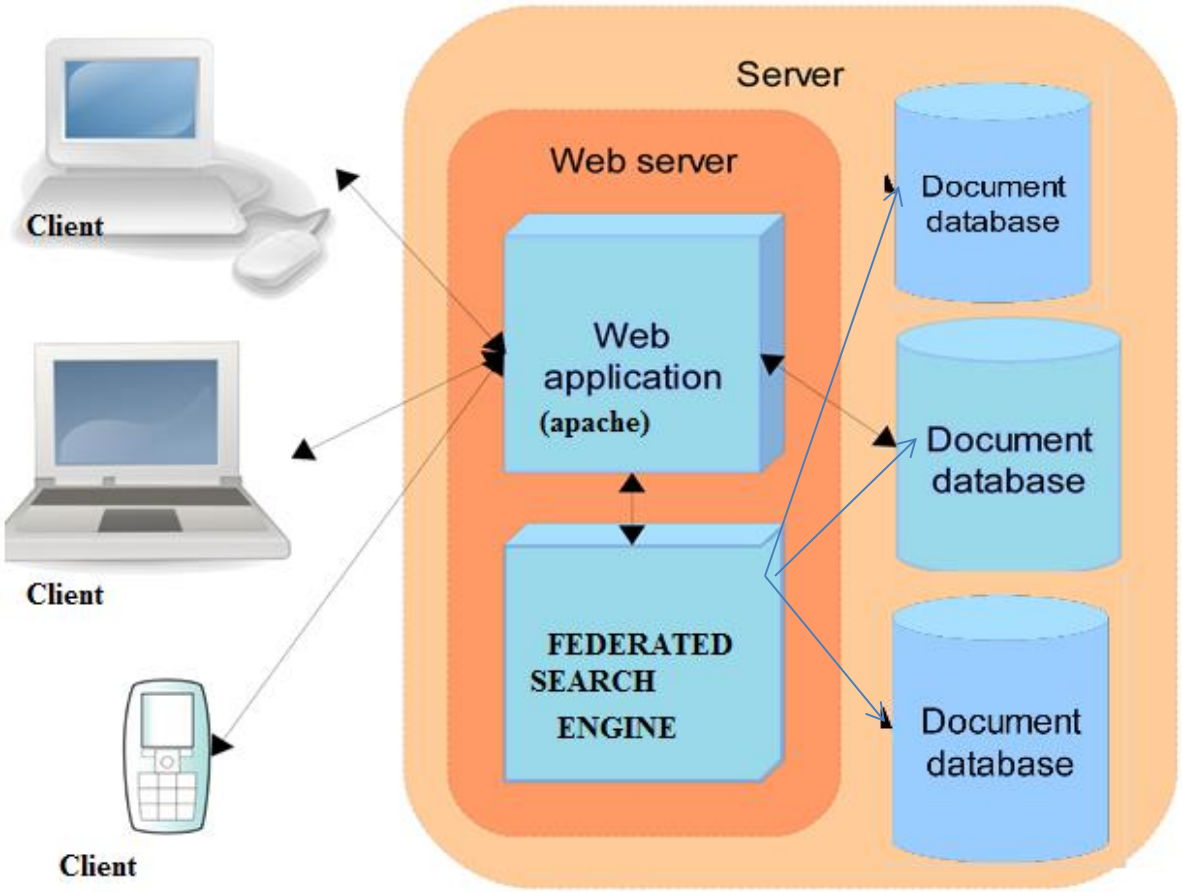

Figure 5-1 project architecture

In the scenario above, the federated search engine runs alongside another application in a Web server. For this project the web server would provide a user interface used for inserting user queries, and a way to make search the research collection in the databases. The collection of researches items would be kept in the databases.

The application is developed through the following steps:

- Define a *schema*. The schema tells the system about the contents of documents it will be indexing. In this project, the schema would define fields for the tables in each database like title, author, subject, and the contents of document and so on.
- Deploy the system to the application server.
- Feed system the document for which the users will search.
- Expose search functionality in the application.

5.4 Features and Processes of the Project

The application enables you to deliver search results from multiple searchable content providers, simultaneously, via one search query. Users are then able to view search results returned from all of the different content sources in a single, integrated page. This means they no longer need to consult each information resource individually.

.The application provides aggregation, ranking and de-duplication features. Aggregation is the process of combining search results from the different sources in some helpful way. A federated search engine might present all of the results from one source then, beneath those results, present the results from the next source, and so on. Aggregation may incorporate sorting (e.g., by date, title, or author), or it may involve ranking, also known as relevance ranking. A researcher searching a couple of dozen sources via a federated search engine usually wants to know which results are most relevant to his or her search from among all of the sources. Relevance ranking compares results from all sources against one another and displays the results in order.

The system submits the user query to a number of content sources, and then combines the results that are returned to one ranked list. It provides a single interface that has a place to enter a user query. The user query might be a title of a research, author name, a key word of a research or any term from the content of the research.

The result is a merged from all databases and they are ranked based on the term similarity with the document. The *figure 5.2* below shows the search interface of the system.

Figure 5-2 Search interface of the system

In addition to a query box, the search interface by default displays the whole document indexed in the system with their title, author and description. And the users also have a chance to see those documents in the system. Federated search occurs live and results are current and there is no stale content. That means it is scheduled that new contents in the databases are imported to the system every 10 minutes

The other feature of this search engine is spell checking feature. The Spellcheck component is designed to provide inline query suggestions based on other, similar terms. The basis for these suggestions can be terms in a field in the search engine application.

This feature provides a term suggestion when a user enters any query term, it gives a suggestion by asking did you mean plus term suggested by the search engine. The *figure 5.3* below shows the term suggestion features of a search engine.

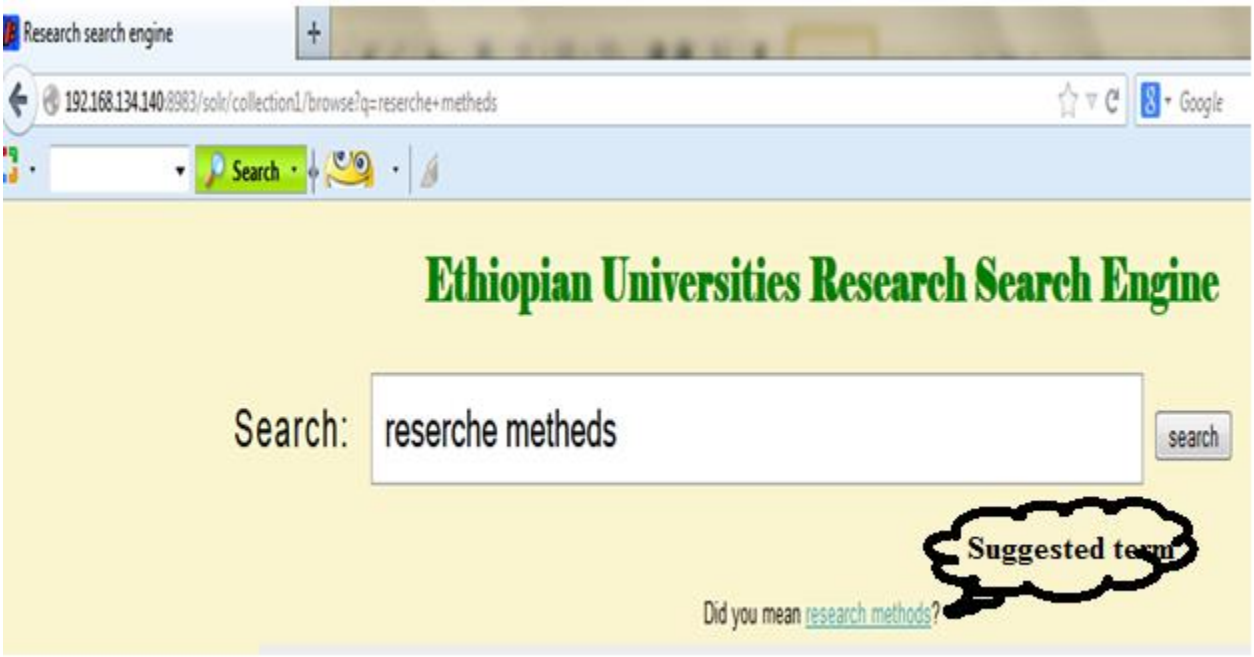

Figure 5-3 Spell checking feature interface of the system

In the above figure, the user enters a term reserche metheds and the system suggests to the correct term research methods by asking *Did you mean research methods?*, *If* the user accepts the suggested term, the system displays the result page that contains a documents with a term.

The result page of the system contains the title of the research, Author, and description of the research that contains a query term. Additionally, the result page includes a sample text from the research document and a link to full-text research. *Figure 5.4* below shows the result page of the system.

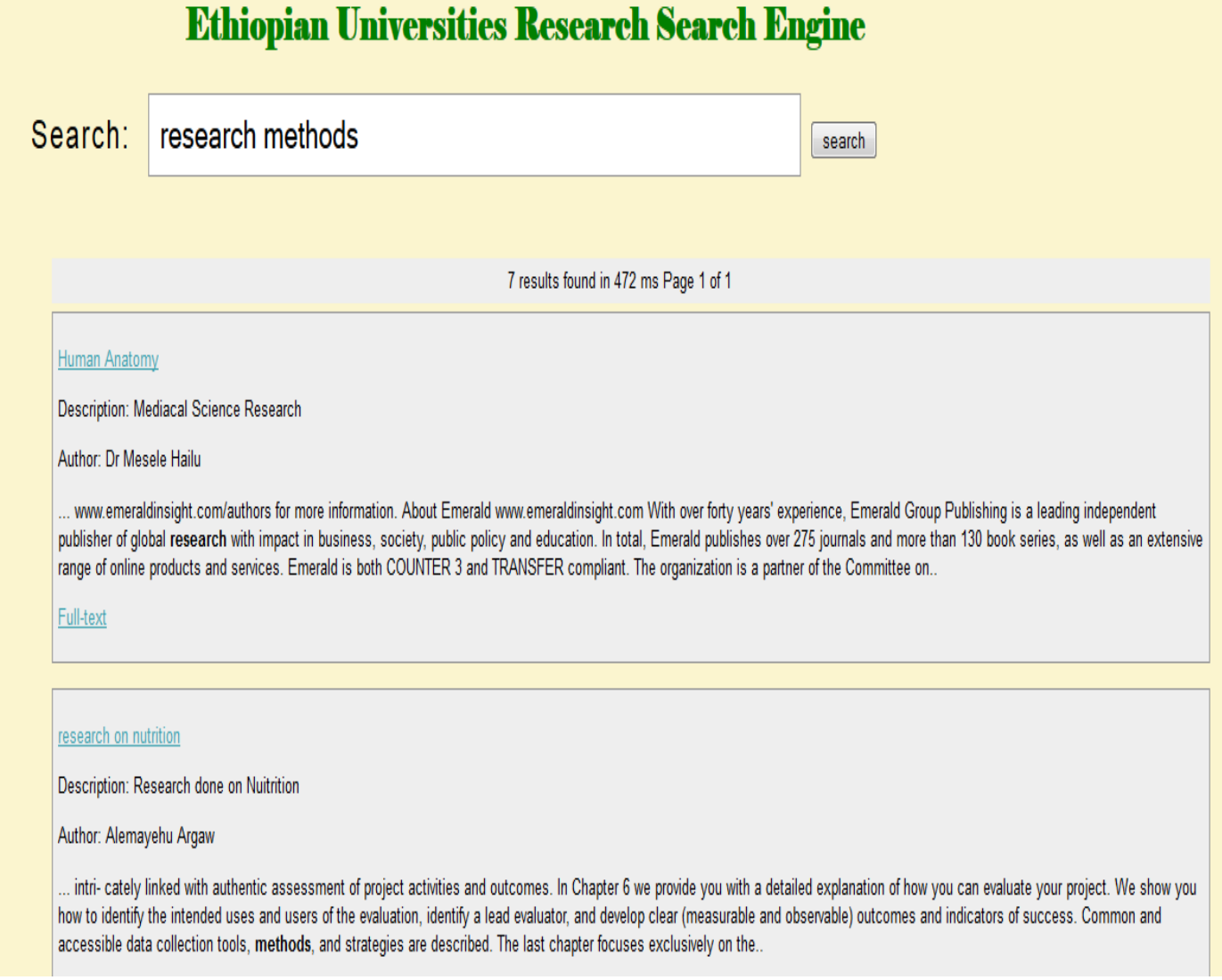

Figure 5-4 Search result interface of the system

Once the users get this result page with necessary information about the document, they can access to a full document by just clicking to full-text link. Additionally, content is combined from different information sources saving efforts of searching sources one at a time.

The process of this federated search engine application consists of:

- Indexing. This federated search engine application uses Inverted Index method for indexing words in the document. When a user enters a query into a system, the engine examines its index and provides a listing of best matching resources according to its criteria, with a short summary containing the document's title, author, some parts of the text and a link to full-text of the resource. 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.
- Tokenization and filtering: Tokenization break field data into lexical units, or *tokens* and filtering examine a stream of tokens and keep them, transform or discard them, or create new ones.
- Transforming a query and broadcasting it to a group of disparate databases with the appropriate syntax,
- Search. The usefulness of a search engine depends on the relevance of the result set it gives back. While there may be lots of resources that include a particular word or phrase, some resources may be more relevant, popular, or authoritative than others. This system employs Inverse document frequency (idf) method to rank the results to provide the "best" results first.
- Providing a merged result set to a user and,
- Scheduled data-import from different databases every 10 minutes to check, import and index new documents inserted to those databases.

5.5 Challenges in Practical Implementation of the Project

When this federated search application is performed against secure data sources, the users' credentials must be passed on to the search engine, so that appropriate security is maintained. If the user has different login credentials for different systems, there must be a means to map their login ID to each search engine's security domain. So, in order to access the contents in each database the credential of each database is needed.

That means Verification, authentication, and certification can be difficult for the federated search. Since this federated search engine don't hold the data locally, meaning the engines perform the search and send the results back, this federated search engine must be able to access multiple, password-protected databases behind the scenes, all at one time, and show users their results in one easy-to-read interface. The challenge for this federated search is to ensure that only licensed users can access databases in an appropriate manner, as specified by their license. This may require a library or a university of a local research database to set up multiple areas where only certain licensed users can access a federated search.

CHAPTER SIX CONCLUSION AND RECOMMENDATION

In the electronic information environment one of the responses to the problem of bringing large amounts of information together has been for libraries to introduce portals. Currently, only few of Ethiopian universities have online research databases. Students and research communities of the universities have a high need for a use of online researches in their university in order to using them as reference for their study.

Federated search is a technique for searching multiple text collections simultaneously. This study shows that there is a big demand and need of federated search engine application that fills the gap of research information sharing problem between Ethiopian universities. Additionally, this research project studied the usage of online research databases, the research information sharing problem and the need of federated search engine application in Ethiopian universities. Based on the results of the study, federated search engine application prototype is developed in order to simultaneously access multiple research repositories or databases in every university in Ethiopia. The system has one search interface that contains a query box to enter a search term. And the spell check feature that check user query term and suggest a term. The result page of the system provides access to results that meet user query and a link to full research document in the databases.

Finally, every Ethiopian universities have to develop their own research databases in order to make accessible online the research works in their university and gain benefits from access to other universities research findings. The federal ministry of education should have to take the responsibility to implement this federated search engine application to create research information sharing environment between Ethiopian universities.

REFERENCES

Apache solr reference guide, (2013), http://lucene.apache.org/solr/.

Baeza-Yates, R. &Ribeiro-Neto, B. (1999). Modern information retrieval.ACM Press / Addison Wesley.

Bergman, M. (2001). The deep web: surfacing the hidden value. http://www.completeplanet.com/Tutorials/DeepWeb/index.asp. BrightPlanet.

Bonetti, L., Ceppi, S. &Gatti, N. (2008). Designing a Revenue Mechanism for federated Search Engines

Callan, J., Croft, W. B. & Harding, S. M. (1992). The INQUERY retrieval system.In *Proceedings of theThird International Conference on Database and Expert Systems Applications*, (pp. 78-83).ACM.

Callan, J., Croft, W. B. &Broglio, J. (1995). TREC and TIPSTER experiments with INQUERY. *InformationProcessing and Management*, 31(3). (pp. 327-343).

Callan, J., Lu, Z. & Croft, W. B. (1995). Searching distributed collection with inference networks. In*Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development inInformation Retrieval.*ACM.

Callan, J., Connell, M. & Du, A. (1999). Automatic discovery of language models for text databases. In*Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*.ACM.

Callan, J. (2000). Distributed information retrieval.In W.B. Croft, editor, *Advances in Information Retrieval*.(pp. 127-150). Kluwer Academic Publishers.

Callan, J. & Connell, M. (2001). Query-based sampling of text databases.*ACM Transactions on InformationSystems, 19(2)* (pp. 97-130). ACM.

Chen, Xiotin.(2006). MetaLib, WebFeat, and Google: The strengths and weaknesses of federatedsearch engines compared with Google. Online Information Review, 30(4),.p 413-427. Available at http://www.emeraldinsight.com/Insight/ViewContentServlet?Filename= Published/EmeraldFullTextArticle/Articles/2640300406.html

Craswell, N., Hawking, D., &Thistlewaite, P. (1999). Merging results from isolated search engines. In*Proceedings of the 10th Australasian Database Conference.*(pp. 189-200).

Craswell, N. (2000). Methods for distributed information retrieval.Ph. D. thesis, Department of ComputerScience, The Australian Nation University.

Curtis, Marie, A., Gorner & Daniel G. (2005). Why Federated Search? Knowledge Quest, 33 (3),January/February 2005. p. 37.

Federated Search Engine 2001-2003.http://www.ala.org/ala/alctscontent/alctspubsbucket/ webpublications/cataloging/researchtopics/federated.cfm

French, J. C., Powell, A. L., Callan, J., Viles, C. L., Emmitt, T., Prey, K. J., &Mou, Y. (1999). Comparing theperformance of database selection algorithms.In *Proceedings of the 22nd Annual International ACM SIGIRConference on Research and Development in Information Retrieval*.ACM.

Fryer, D. (2004). Federated search engines: federated searching aggregates multiple channels of information into a single searchable point, Online, 28 (2), March-April,2004. p.16.

Gravano, L., & Garcia-Molina, H. (1995). Generalizing GlOSS to vector-space databases and brokerhierarchies. In *Proc. of 21th International Conference on Very Large Data Bases (VLDB'95*), pp 78-89.

Gravano, L., Chang, C., García-Molina, H., &Paepcke, A. (1997). STARTS: Stanford proposal for internetmeta-searching. In *Proceedings of the ACM-SIGMOD International Conference on Management of Data.*ACM.

Kirsch, S. T. (1997). Document retrieval over networks wherein ranking and relevance scores are computedat the client for multiple database documents. U.S. Patent 5,659,732.

Koster.(1994). ALIWEB, Archie-like indexing in the web.Computer Networks and ISDN Systems, 27(2):175{182, 1994. ISSN 1389-1286.

Kumar, S., Sanaman, G. & Rai, N. (2008). .Federated Search: New Option for Libraries in the Digital Era.

Liu. K. L., Yu. C. &Meng.W., Santos.A. & Zhang.C. (2001).Discovering the representative of a searchengine.In *Proceedings of 10th ACM International Conference on Information and Knowledge Management (CIKM).*ACM.

Raghavan and H. Garcia-Molina. (2001), Crawling the hidden web. In Aperset alpages 129 - 138. ISBN 1-55860-804-4.

Roy, T. (2001). Digital libraries: cross database search. Library Journal.Availableat http://libraryjournal.reviewsnews.com/index.asp?layout=articlePrint&articleID=CA170458

SeCo.(2006). http://www.search-computing.it/s.

Shokouhi, M., Luo, S. (2009). Now the essence of knowledge: Federated search.

Si, L. &Callan, J. (2002a). Using sampled data and regression to merge search engine results. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development inInformation Retrieval.*ACM.

Si, L., Jin, R., Callan, J. & Ogilvie, P. (2002b). A language model framework for resource selection andresults merging.In *Proceedings of the 11th International Conference on Information and KnowledgeManagement.*ACM.

Si, L. & Callan., J. (2003a). Relevant document distribution estimation method for resource selection.In*Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development inInformation Retrieval.*ACM.

Si, L. & Callan, J. (2003b). A Semi-Supervised learning method to merge search engine results. *ACMTransactions on Information Systems*, 21(4). (pp. 457-491). ACM.

Si, L. & Callan, J. (2003c). Distributed information retrieval with skewed database size distributions.*InProceedings of the NSF's National Conference on Digital Government Research (dg.o2003.)*

Spink, B. Jansen, C. Blakely, and S. Koshman.(2006). A study of resultsoverlap and uniqueness among major web search engines.Information Processing and Management, 42(5):1379{1391, 2006. ISSN0306-4573.

Voorhees, E., Gupta, N. K., & Johnson-Laird, B. (1995). Learning collection fusion strategies.In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development inInformation Retrieval*.ACM.

Xu, J. & Croft., W. B. (1999). Cluster-based language models for distributed retrieval. In *Proceedings of the22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.ACM.

Yuwono, B. & Lee, D. L. (1997). Server ranking for distributed text retrieval systems on the Internet. *InProceedings of the 5th Annual International Conference on Database Systems for Advanced Applications.*(pp. 41-49). World Scientific Press.

Annex 1: Questionnaire for Postgraduate Students

Purpose of the Survey

Federated search engine application is an application that can provide parallel search over multiple collections on multiple databases. This survey provides an opportunity to share your thoughts on what you feel on the federated search engine application tool is needed to ensure that there is need of application in order to develop the federated search engine application used to bridge the research databases of Ethiopian universities to create research information sharing environment.

- You do not have to fill out this survey if you do not want to. However, everyone's views are important.
- This questionnaire should be filled by post graduate students.
- If you have any question or unclear question, please don't hesitate to ask the data collector.
- If you have any comments or suggestion on the research area or any issues with this survey contact the researcher with this email address dula.boru@ju.edu.et.

Thank you for your help by providing your thought and feeling on the need of federated search engine application that help to develop the federated search application Ethiopian universities

Instructions

- Please read each question carefully and answer as accurately as you can.
- There are two types of questions. One type requires you to given space beside the question. The other type of question requires you to place either a √or an **X** in the box beside your response.

1. University: _____

2. Job responsibility in your university: _____

3. Age:

❒<18 yrs ❒ 26-35 yrs

❒ 19-25 yrs ❒ 36-50 yrs

❒>51 yrs

4. Gender:

❒Male

❒Female

5. Occupation:

❒Undergraduate student ❒ Postgraduate Student

❒ Instructor ❒ Administrative Staff

❒ Other _____

6. Computer Usage Frequency in hrs per day:

❒<1 hr/day ❒ 3-5 hrs/day

❒1-2 hrs/day ❒>5 hrs/day

Other specify _____

7. Internet Usage Frequency in hrs per day:

❒<1 hr/day ❒3-5 hrs/day

❒1-2 hrs/day ❒>5 hrs/day

Other specify_____

8. What is your experience & usage frequency using search engine?

❒<1 hr/day ❒3-5 hrs/day

❒1-2 hrs/day ❒>5 hrs/day

Other specify_____

9. Are you doing a research?

❒Yes.

❒No.

10. Where do you search for research information?

❒University Research Database

❒Department and/or college/faculty

❒ Library

❒Web Search Engines (Internet)

❒Others_____

11. Which search engine do you frequently use in order to search for information about your research?

❒ Google ❒Excite ❒Others, Specify_____

❒ Yahoo ❒Lycos

❒ AltaVista ❒MSN Search

12. Do you Access other universities Research information such as student and staff scholarly papers, reviews, project works and documents?

❒Yes.

❒No.

13. If Yes, How?

_____

_____

14. If No, Why?

_____

_____

15. Does other universities research information is necessary for your study?

❒Yes

❒No

16. What are benefits of accessing other universities research information databases?

_____

_____

_____

17. Do you think there is a problem of research information sharing between Ethiopian universities?

❒Yes

❒ No

18. Do you think there is a need of application used to access other universities database?

❒Yes

❒ No

19. Do you think there is a need of federated search engine application that helps you to search across research databases of all Ethiopian universities?

❒Very likely ❒ Unlikely

❒Likely ❒ Very Unlikely

❒ Neutral

20. How will you use the federated search engine application if it is developed and implemented by universities in Ethiopia?

❒Very extensively

❒ Extensively

❒ Neutral

❒ Rarely

❒ None

21. In your opinion what features should the application should include for you to use it efficiently and effectively? Can you list them?

_____

_____

_____

_____

Annex2: Questionnaire for Academic Staffs

Purpose of the Survey

Federated search engine application is an application that can provide parallel search over multiple collections on multiple databases. This survey provides an opportunity to share your thoughts on what you feel on the federated search engine application tool is needed to ensure that there is need of application in order to develop the federated search engine application used to bridge the research databases of Ethiopian universities to create research information sharing environment.

- You do not have to fill out this survey if you do not want to. However, everyone's views are important.
- This questionnaire should be filled by academic staffs.
- If you have any question or unclear question, please don't hesitate to ask the data collector.
- If you have any comments or suggestion on the research area or any issues with this survey contact the researcher with this email address dula.boru@ju.edu.et.

Thank you for your help by providing your thought and feeling on the need of federated search engine application that help to develop the federated search application Ethiopian universities

Instructions

- Please read each question carefully and answer as accurately as you can.
- There are two types of questions. One type requires you to given space beside the question. The other type of question requires you to place either a √or an **X** in the box beside your response.

1. University: _____

2. Job responsibility in your university: _____

3. Age:

❒<18 yrs ❒ 26-35 yrs

❒ 19-25 yrs ❒ 36-50 yrs

❒>51 yrs

4. Gender:

❒Male

❒Female

5. Computer Usage Frequency in hrs per day:

❒<1 hr/day ❒ 3-5 hrs/day

❒1-2 hrs/day ❒>5 hrs/day

Other specify _____

6. Internet Usage Frequency in hrs per day:

❒<1 hr/day ❒3-5 hrs/day

❒1-2 hrs/day ❒>5 hrs/day

Other specify_____

7. What is your experience & usage frequency using search engine?

❒<1 hr/day ❒3-5 hrs/day

❒1-2 hrs/day ❒>5 hrs/day

Other specify_____

8. Do you have research experience?

❒Yes.

❒No.

9. Where do you search for research information?

❒University Research Database

❒Department and/or college/faculty

❒ Library

❒Web Search Engines (Internet)

❒Others_____

10. Which search engine do you frequently use in order to search for information about your research?

❒ Google ❒Excite ❒Others, Specify_____

❒ Yahoo ❒Lycos

❒ AltaVista ❒MSN Search

11. Do you advise student research?

❒Yes.

❒No.

12. If yes, is the student copying their research from other universities?

❒Yes.

❒ Partly

❒No.

13. Is there a method or an application to check either the research is copied from other university or it is a new work?

❒Yes.

❒No.

14. If yes. How do you check either the student research is copied from other university?_____

_____

_____

15. Do you think a lack of application to access and other universities research information have an impact on student research?

❒Yes.

❒No.

16. If yes. Can you mention those impacts?-

_____

_____

_____

17. Does other universities research information is necessary for students and also for your study?

❒Yes

❒No

18. What are benefits of accessing other universities research information databases?

_____-

_____

_____

19. Do you think there is a problem of research information sharing between Ethiopian universities?

❒Yes

❒ No

20. Do you think there is a need of application used to access other universities database?

❒Yes

❒ No

21. Do you think there is a need of federated search engine application that helps you to search across research databases of all Ethiopian universities?

❒Very likely ❒ Unlikely

❒Likely ❒ Very Unlikely

❒ Neutral

22. How will you use the federated search engine application if it is developed and implemented by universities in Ethiopia?

❒Very extensively

❒ Extensively

❒ Neutral

❒ Rarely

❒ None

23. In your opinion what features should the application should include for you to use it efficiently and effectively? Can you list them?

_____

_____

_____

_____