

Mining Linked data for Search Optimization

By

Md. Al-amin Tusar (ID: 2012-3-60-019)

Safwan Mohsin (ID: 2012-3-60-035)



Department of Computer Science and Engineering

East West University

Supervised By

Md. Sarwar Kamal, Senior Lecturer

Department of Computer Science and Engineering

East West University

December, 2016

Mining Linked data for Search Optimization

Submitted by

Md. Al-amin Tusar

ID: 2012-3-60-019

Safwan Mohsin

ID: 2012-3-60-035

**This Project is submitted for partial fulfillment of the
requirements for the degree of Bachelors of Science (B.Sc.) in
computer Science and Engineering**

In the

Faculty of Science and Engineering

Department of Computer Science and Engineering

East West University

Fall 2016

Declaration

We hereby declare that the project titled “Mining linked data for search optimization” is record of original work done by us. To the best awareness of us, it has not been collected from any source without the due acknowledgement and permission. This is submitted for the partial requirement of the degree of Bachelor of Science (B.Sc.) in Computer Science and Engineering, East West University. This project report is not submitted to any other university or institution for the award of any degree or fellowship or published any time before.

Md. Al-amin Tusar

ID - 2012-3-60-019

Safwan Mohsin

ID - 2012-3-60-035

Letter of Acceptance

The project entitled “**Mining linked data for search optimization**” submitted by Safwan Mohsin[2012-3-60-035] and Md. Al-Amin Tusar[2012-3-60-019], to the Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh in the Semester of Fall 2016 has been accepted as satisfactory for partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering.

Md. Sarwar Kamal

Supervisor

Senior Lecturer

Department of Computer Science and Engineering

East West University

Dhaka-1212, Bangladesh

Dr. Mozammel Huq Azad Khan

Professor and Chairperson

Department of Computer Science and Engineering

East West University

Dhaka-1212, Bangladesh

Abstract

Searching information in internet is a natural activity in current situation. People are searching in web for all sorts of information, what they need in their daily life. The information in the web is spread among lots of data. Some information's are and frequently searched by the user. The searched information of different users may be matched. Frequent data indicates that, those data might be related, which are searched in a particular time.

Those related (raw linked data) might be used in future to give some interesting information or to optimize the search of a user. Searching data in broad (web) area is quite difficult process to retrieve information. Searching in the web is also matter of cost and need good throughput if the traditional search engines are followed. In short area of data searching, it is quite effective to fetch specific information. It is a challenging issue to reduce the search area and categorize the reduced data in such a way that the important data is not lost. In this project we have used Apriori algorithm to mine a huge set of data to generate some optimized linked data which decreases the search area of search engine and also gives an optimized result to the users who are searching the information. We have proposed a model for the implementation with a tool and also demonstrated how the distributed computing can be used to optimize huge data if required.

Acknowledgement

Firstly, we would like to express my indebtedness appreciation to our supervisor **Md. Sawrar Kamal**, Senior lecturer of Department of Computer Science and Engineering. His constant guidance and advice played the vital role in making the execution of the report. He always gave me his suggestions that were crucial in making this report as flawless as possible. We would like to also express our gratitude to **Dr. Shamim H. Ripon**, Associate Professor of Department Computer Science and Engineering, who also helped us to complete this project with lots of encouragement and support. We are also thankful to our parents, other faculties of the department and friends for the support and inspiration regarding our project.

Table of Contents

Chapter 1	1
1. Introduction	1
1.1 Motivation	2
1.2 Objectives	2
1.3 Contribution.....	3
1.4 Outline	3
Chapter 2	4
2. Background	4
2.1 Data Mining.....	4
2.2 Background of data mining	4
2.3 Why data mining	5
2.4 Web mining	6
2.5 Web usage mining	7
2.6 Searching and mining	8
2.8 Current situation of search.....	9
2.9 Related Works	10
Chapter 3	15
3. Proposed Model	15
3.1 Overview of the system.....	15
3.2 Data collection.....	17
3.3 Organizing data	17
3.4 Mining data.....	18
3.5 Optimized data.....	19
Chapter 4	20
4. Implementation.....	20
4.1 Tools description	20

4.1.1 Languages	20
4.1.2 Database management	22
4.1.3 Visualization	23
4.1.4 Support for Running Our Program	24
4.2 User Manual	25
4.2.1 Login Procedure.....	25
4.2.2 Searching Procedure	26
4.3 Client Side	27
4.3.1 Performable Search Type	27
4.3.2 Input Acquisition	31
4.3.3 Differentiating Between Users.....	32
4.4 Server Side.....	33
4.4.1 Session	33
4.4.2 Collecting Raw linked data.....	33
4.4.3 Effect of Mining Algorithm	35
4.5 Outcome	39
Chapter 5	40
5. Result Analysis.....	41
Chapter 6	44
6. Conclusion and Future Work.....	44
6.1 Conclusion.....	44
6.2 Future Work.....	45
Bibliography	46
Appendix.....	47
A.1 Home Page	47
A.2 Collecting User IP	48
A.3 Searching Mechanisms	49
A.4 Adding New Websites to Database	51
A.5 Information Grabbing To Store with newly collected site	54

A.6 Data Acquisition for Processing	55
A.7 Conversion of Temporary User Data to Raw Linked Data and Logging out a Session	57
A.8 Optimizing Linked data by Mining	59
A.9 Providing Benefit to User	67

LIST OF FIGURES

1. Figure 2.1: Standard data mining process.....	5
2. Figure 2.2: Google search statistics in several years.	10
3. Figure 3.1: Proposed model for optimizing linked data	16
4. Figure 3.2: Data Collection from user.	17
5. Figure 3.3: Implementation of Apriori algorithm for mining.	18
6. Figure 3.4: Optimized data in different degree of Apriori.....	19
7. Figure 4.1: Sublime Text Editor Coding Window.....	23
8. Figure 4.2: Xampp Server Solution	24
9. Figure 4.3: Visualization of program in a browser.....	24
10. Figure 4.4: Active Server and Database.	25
11. Figure 4.5: Login menu.....	25
12. Figure 4.6: Performing Search.....	26
13. Figure 4.7: The Array index and their values	29
14. Figure 4.8: ‘Input Acquisition’	31
15. Figure 4.9: ‘website’ table	32
16. Figure 4.10: ‘temptable’ table.....	33
17. Figure 4.11: ‘transaction’ table.....	34
18. Figure 4.12: Workflow of Apriori algorithm.....	35
19. Figure 4.13: 1 st degree Apriori applied on ‘transaction’ table	36
20. Figure 4.14: 2 nd degree Apriori applied on ‘transaction’ table.	37
21. Figure 4.15: 3 rd degree Apriori applied on ‘transaction’ table.....	38
22. Figure 4.16: User getting benefit of optimization.....	39
23. Figure 5.1: Comparison between optimized and un-optimized data	42

LIST OF TABLES

1. Table 3.1: Organizing data in item sets	18
2. Table 5.1: Cell traversal data computation	43

Chapter 1

Introduction

Searching for information in web is now part and parcel of our everyday life. Huge collection of data is in virtual world and lots of them are in form of unusable log data also known as data tomb. All types of information is useful in different perspective. Retrieving specific information from this huge data or even from data tomb is quite difficult task. Generally search engines are used to find out any information which is spread in the websites. The desired website addresses are the actual information link to the desired data. These web addresses leads us to the website which have the information structured in a webpage. [1]

Search optimization is a technique to reduce the stress or the runtime of any reply of a query such as, the user gets his result instantly and also some linked data with his/her result which may also useful to him. The optimization is mainly focused on frequency of data. It reduces the burden of analyzing huge number of data because the searching area is reduced. The optimization model is applied after users search for any input and the data is stored for further processing, so that after model being applied user can receive quicker response for future search. Every time the algorithm will run the optimized result will be updated. [9]

Frequent data are the now revolutionizing the conceptual foundations of many fields. When it is distributed into multithreading the processing time will be reduced. In future, some improved pattern or knowledge might be generated after implementing the search optimization. Web crawler, and RDF and mapping can be used to improve the collection technique. For any specific research or implementation, the input data may be categorized in the preferable way of research/implementation. The runtime of the program may be reduced more if the distributed computing is applied. [9]

1.1 Motivation

In modern world we have lots of data in the internet and the amount of data there is huge. So as users want quicker response we need to retrieve any data as quick as possible. In this area of context, several information retrieval technologies have been developed to assist users to fulfill their searching needs on web. The most commonly used search engines by users are Google, Yahoo, Amazon, Bing, Ask, Search and so on. The search engines allow users to find relevant resources needed by setting up their query patterns and reviewing a list of URLs. It is very good to have such advanced search engines. But most of the time, for any specific task we need to search for many websites. And for a particular job it could be related. A person who wants to set up a fish pond on his backyard may want to search for pond building techniques, suitable fish for that environment, and guide to buy pond accessories. Search engine does help him finding all his desired result but to do so he need to search thrice. What if only 1 search can provide him everything he needs and at the same time in optimized way?

Suggestion can be generated for a person from other user's data if they were stored somewhere. But to do so, a system has to go through all the data it collected so far, which is insanely big to think about. In this paper we propose a model which can provide quicker response to user by reducing the number of cells to be explored when searched. And to do that we used frequent item set mining algorithm (Apriori).

1.2 Objectives

The specific object of the project as follows:

- Collection of data form users and store in a specific structure.
- Mining the structured data according to their frequency.
- Find the optimized data for further use.

1.3 Contribution

Contribution in the project is as follows:

We have used html anchor link and JavaScript function for data collection and MySQL database server to store the information in database.

The Apriori algorithm has been implemented for frequent data searching. Apriori generates optimized set of data to reduce the searching.

For the Apriori algorithm implementation we have also used MySQL query to handle the database tables of MySQL server.

1.4 Outline

Chapter 1: This chapter represents about the motivation to work, specify the objectives and then the contribution we have made.

Chapter 2: Brief discussion of data mining, how searching and data mining is related. And current outlook of search, recent works and researches of search optimization and some of their portion which are related to our project.

Chapter 3: Proposed model of our project. Here the entire model is shown and how will the work step by step.

Chapter 4: Implementation and flow diagram of our project in terms of client and server side.

Chapter 5: Result analysis of our project. How the output of the tool is giving for different numbers of input.

Chapter 6: The future work, the improvement in our work in long term and the summarization of our work.

Chapter 2

Background

2.1 Data Mining

Data mining is the process of discovering or obtaining interesting (non-trivial, implicit, previously unknown and potentially useful) patterns and knowledge from huge amounts of data. It is actually used for knowledge discovery, which is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining is a process used by companies to turn raw data into useful information, by using software to look for patterns in large batches of data,

Data mining consists of five major elements:

- 1) Extract, transform, and load transaction data onto the data warehouse system.
- 2) Store and manage the data in a multidimensional database system.
- 3) Provide data access to business analysts and information technology professionals.
- 4) Analyze the data by application software.
- 5) Present the data in a useful format, such as a graph or table. [2][14]

2.2 Background of Data Mining

The manual extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has dramatically increased data collection, storage, and manipulation ability. As data sets have grown in size and complexity, direct "hands-on" data analysis has increasingly been augmented with indirect, automated data processing, aided by other discoveries in computer science, such

as neural networks, cluster analysis, genetic algorithms (1950s), decision trees and decision rules (1960s), and support vector machines (1990s). Data mining is the process of applying these methods with the intention of uncovering hidden patterns in large data sets. It bridges the gap from applied statistics and artificial intelligence (which usually provide the mathematical background) to database management by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery algorithms more efficiently, allowing such methods to be applied to ever larger data sets.[11]

2.3 Why Data Mining

- Gives new information
- Informs about the current trend of the subject.
- Helps to figure out the sudden change in data.
- Gives the idea for the solution of errors in data.[15]

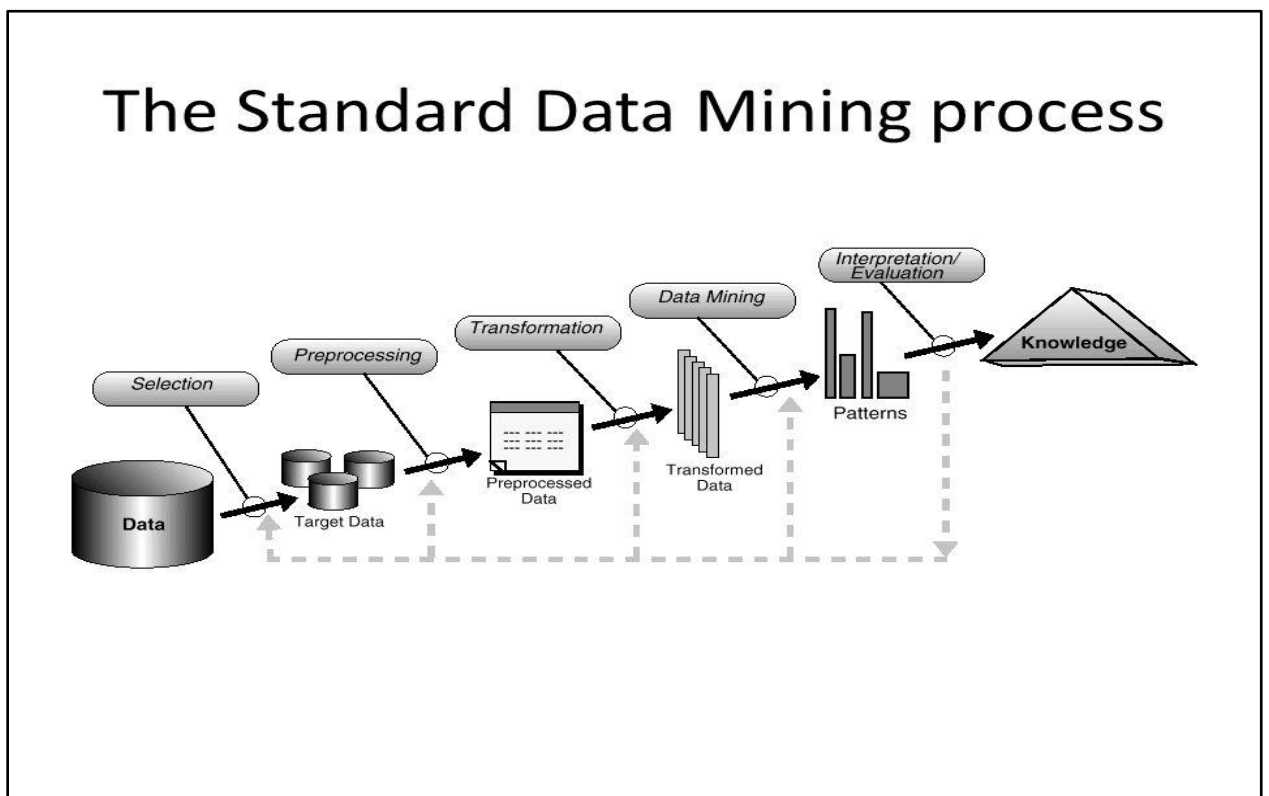


Figure2.1: Standard data mining process

2.4 Web Data Mining

The term Web Data Mining is a technique used to crawl through various web resources to collect required information, which enables an individual or a company to promote business, understanding marketing dynamics, new promotions floating on the Internet, etc. There is a growing trend among companies, organizations and individuals alike to gather information through web data mining to utilize that information in their best interest.[11]

Data Mining is done through various types of data mining software. These can be simple data mining software or highly specific for detailed and extensive tasks that will be sifting through more information to pick out finer bits of information.[9]

Before this data mining software came into being, different businesses used to collect information from recorded data sources. But the bulk of this information is too much too daunting and time consuming to gather by going through all the records, therefore the approach of computer based data mining came into being and has gained huge popularity to now become a necessity for the survival of most businesses.[11][12]

The collected information is used to gain more knowledge and based on the findings and analysis of the information make predictions as to what would be the best choice and the right approach to move toward on a particular issue. Web data mining is not only focused to gain business information but is also used by various organizational departments to make the right predictions and decisions for things like business development, work flow, production processes and more by going through the business models derived from the data mining. [9][12]

The data mining technology is going through a huge evolution and new and better techniques are made available all the time to gather whatever information is required. Web data mining technology is opening avenues on not just gathering data but it is also raising a lot of concerns related to data security. There is loads of personal information available on the internet and web data mining had helped to keep the idea of the need to secure that information at the forefront.[11][12]

2.5 Web Usage Mining

Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. [9]

Web usage mining itself can be classified further depending on the kind of usage data considered:

Web Server Data: The user logs are collected by the Web server. Typical data includes IP address, page reference and access time.

Application Server Data: Commercial application servers have significant features to enable e-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs.

Application Level Data: New kinds of events can be defined in an application, and logging can be turned on for them thus generating histories of these specially defined events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the categories above.

Studies related to work are concerned with two areas: constraint-based data mining algorithms applied in Web Usage Mining and developed software tools (systems). Costa and Seco demonstrated that web log mining can be used to extract semantic information (hyponymy relationships in particular) about the user and a given community. [11][12]

Pros

Web usage mining essentially has many advantages which makes this technology attractive to corporations including the government agencies. This technology has enabled e-commerce to do personalized marketing, which eventually results in higher trade volumes. Government agencies are using this technology to classify threats and fight against terrorism. The predicting capability of mining applications can benefit society by identifying criminal activities. The companies can establish better customer relationship by giving them exactly what they need. Companies can understand the needs of the customer better and they can react to customer needs faster. The companies can find, attract and retain customers; they can save on production costs by utilizing the acquired insight of customer

requirements. They can increase profitability by target pricing based on the profiles created. They can even find the customer who might default to a competitor the company will try to retain the customer by providing promotional offers to the specific customer, thus reducing the risk of losing a customer or customers.[12]

Cons

Web usage mining by itself does not create issues, but this technology when used on data of personal nature might cause concerns. The most criticized ethical issue involving web usage mining is the invasion of privacy. Privacy is considered lost when information concerning an individual is obtained, used, or disseminated, especially if this occurs without their knowledge or consent. The obtained data will be analyzed, and clustered to form profiles; the data will be made anonymous before clustering so that there are no personal profiles. Thus these applications de-individualize the users by judging them by their mouse clicks. De-individualization, can be defined as a tendency of judging and treating people on the basis of group characteristics instead of on their own individual characteristics and merits.[12]

Another important concern is that the companies collecting the data for a specific purpose might use the data for a totally different purpose, and this essentially violates the user's interests.

2.6 Searching and Data Mining:

When a user, searches something in the search engine, Search Engine provides the gateway for most of the users trying to explore the huge information base of web pages. Search engines are programs that search documents for specified keywords on search for information on the World Wide Web and return a list of the documents where the keywords were found. But huge data are searched in the search engines, from which figuring out the desired information is quite difficult. So this huge data is have to be mined such a way that the search engine do not have to search this huge data to answer a user's query. And also can give suggestion to user for the similar query. So the searches of the user have to be analyzed such a way that in future the search engine can give better response and process the query of the user as fast as possible.

2.7 How a Search Engine Works:

There are three basic stages for a search engine: crawling – where content is discovered; indexing, where it is analyzed and stored in huge databases; and retrieval, where a user query fetches a list of relevant pages.

Web crawler:

A Web crawler is an Internet bot which systematically browses the World Wide Web, typically for the purpose of Web indexing (web spidering).

Web search engines and some other sites use Web crawling or spidering software to update their web content or indices of others sites' web content. Web crawlers can copy all the pages they visit for later processing by a search engine which indexes the downloaded pages so the users can search much more efficiently.

Web indexing:

Web indexing (or Internet indexing) refers to various methods for indexing the contents of a website or of the Internet as a whole. Individual websites or intranets may use a back-of-the-book index, while search engines usually use keywords and metadata to provide a more useful vocabulary for Internet or onsite searching. With the increase in the number of periodicals that have articles online, web indexing is also becoming important for periodical websites.

Page rank:

PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites

2.8 Current Situation of Search

According to Google it processes over 40,000 search queries every second on average, which translates to over **3.5 billion searches per day** and **1.2 trillion searches per year** worldwide. There is statistics of searches in several years in the Google search engine.

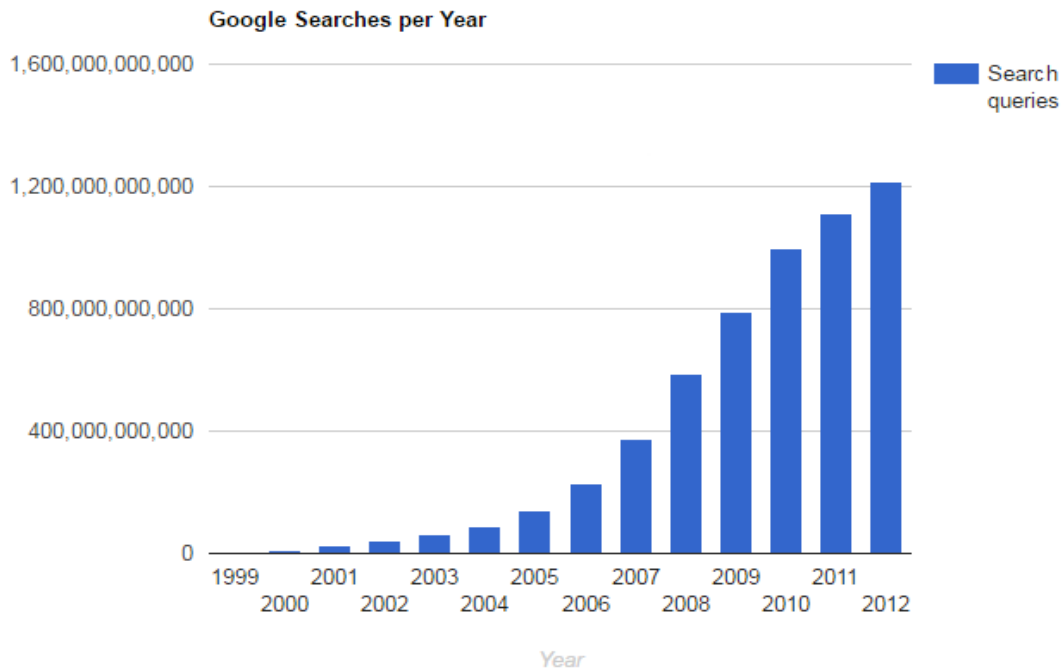


Figure 2.2: Google search statistics in several years.

So it is a very tough process to reply with the search result to this huge number of queries. And now a day there are big amount of information in the web. When a user searches for a data in the web, it is quite challenging to process that huge amount of data and giving back the requested result of the user. It is really a question if these huge sizes of data are actually needed for retrieving search result? Actually it is not needed because if we able to distinguish the data which are mostly searched by the users then it will be helpful to reduce the run time of the processing of the reply of search query and it is also useful to reduce the storage.

2.9 Related Works:

In the comparatively short amount since, there have been several developments that have affected how information technology is talked about and used. The foremost vital of are the expansion of the Internet and also the accessibility of low cost hardware. The technologies for the massive information systems discussed today include the Internet (and intranets and extranets) and Web search. Here are few literature related to our work from which we took help from :->

“ASSOCIATION RULE MINING ALGORITHM FOR WEB SEARCH RESULT OPTIMIZATION”

One of the most challenging goals for the web community is to design search engines that allow users to find resources semantically connected to their queries. The huge size of the web and the vagueness of the most commonly used terms to formulate queries still poses a huge problem to achieve this goal. In this paper a search result optimization method has been proposed. It explores the users' query registered in search engine logs in order to learn how users search and also in order to design algorithms that could improve the precision of the answers suggested to users. The proposed system follows the following steps to accomplish the above tasks:

The module M0 shows the query log of user searching behavior as a text file. In the second step module M1; we create clusters based upon query log's text file obtained from M0. The module M2 is used to apply the association rule algorithms mainly apriori and automated Apriori on the clusters created with M1. The module M2 finally creates an association of page, query, and keyword. The module M3 is used to improve the page rank of pages obtained from M2. The next and final step is using the output of previous all steps to optimize and get the relevant pages for suggested list to the searching users.

Methodology used on Proposed System:

- Query clustering
- Finding page associations
- Finding keyword associations
- Finding query associations
- Rank Improvement

PUBLISHED ON: International Journal of Innovative Science (IJSET),
Engineering & Technology, Vol. 2 Issue 12, and December 2015.

“Search Engine Optimization - Using Data

Mining Approach”

This architecture should be capability of systems working in a Distributed manner. In it all processing should be processed on idle computer. The distributed architecture should not increase Network traffic. All the systems connected in a Network should be operated using some firewalls. The module should be as much as easy to plug and play.

- (1) **Crawler Unit:-** It will crawl a website. It will need to use secondary memory to store the web pages downloaded before analysis. The web pages should be saved on each host system, rather than transferred to the control unit to minimize the network traffic. In crawler unit the technique of Data mining that is cluster may be applied by which similar data elements, similar URL's may be kept as a cluster. Cluster helps us to crawl the different pages. Different types of clustering algorithm may be used to crawl the useful URL's.
- (2) **Control Unit:-** When a crawler request a job or send some data elements the control unit will live on a web server and will be used by it. It will need to save the commands that the user like to be process. This can also be understudied by this example as doc file is saved on server.
- (3) **Messaging System:-** To satisfy the necessity of crawler and control unit the crawler must be able to download and process websites with less transmission with control unit-The crawler unit starting a crawl, sending a message to control unit shows that it is ready to execute a new request. The problem is that control unit outside the network can not initiate communication with component inside but only can send information in response to a request for it. This architecture can be successful for any component where the control unit is on a public access web server. The architecture described here is employed to design a system for the job of analyzing the link structure of web sites. This program had not run quickly enough to consider necessary number of websites and so it has been individually setup and run on a number of computers parallel. One more feature that was built into the crawler was a option of types of checking for duplicate pages to be used in a website crawling.

There are three options-

- (1) Use no page checking thinks that two page with different pages are different pages.
- (2) Use HTML page checking oppose new page which is identical HTML to later retrieved page.
- (3) Use weak HTML Page Checking.

Clustering is a method in which like records are grouped together. This technique is done to give the end user a high level view of what is going on in the data set. Sometimes clustering is performed not so much to keep records together as to make it easier to see when one record sticks out from the rest. Clusters may be created either statistically or by using artificial intelligence methods. Clusters can be analyzed automatically by a program or by using visualization techniques

PUBLISHED ON: International Journal of Application or Innovation in Engineering & Management (IJAIEM) , Vol. 2 Issue 9,September 2013

“Target Advertising via Association Rule Mining (using Apriori)”

Data mining supports different techniques of knowledge extraction for various kinds of predictions like clustering, classification, association rule mining, sequential pattern discovery and analysis. Now data mining play an important role in the field of researches, Business, analysis, Predictions. We use association rule mining technique to develop our application. It help us in describing and analyzing of data and presenting strong rules discovered in databases using different measures of interestingness. Introduced association rules for showing regularities between products in large transaction data For example, the rule {diaper} \Rightarrow {rash cream} It indicates if people are buying diaper than they can also buy rash cream. We are using association rule mining to show the relation between different items. It can find a correlation among products from the analysis of a large set of data. The association rule mining is also known as a market basket analysis. It generally works on to know the customers buying pattern. Market basket analysis shows the

different combination of items in the customer's basket. Many Supermarkets use this technique. It also helps us to introduce new products in market. We are using association rule mining in our proposed approach. We know that internet is a great source of marketing. The webpage's have a specific area for marketing or advertising. Web advertisement is the field of research that always looking for the interest of customer. In our research we want to develop a system that introduces advertising with patterns.

METHODOLOGY USED

- (1) Extract the noun from Statements.
- (2) Store the entities at publisher Database.

PUBLISHED ON: International Journal of Advance Research in Computer Science and Management Studies (IJARCSMS), Volume 2, Issue 5, May 2014

Few more paper worth mentioning:

1. Discovering association rules between items in large databases. (Tomasz Imielinski, Arun Swami)
2. Keeping track of web users browsing behavior (Qiang Yang, TianyiliKe Wang)
3. Webusage mining with the association rule mining to optimize the content of the serve log data.(Viral Kapadia ,NileshPrajapati)
4. The association rules in large databases using modified Apriori (Farah Hanna AL-Zawaidah, Yosef HasanJbara)

Chapter 3

Proposed Model

3.1 Overview of the System

In current world, there is massive amount of data in the web. And a very high percentage of people use web for doing most of their daily works. And to do that they often have to explore the unknown as it's not practically possible to memorize all the websites, a person may require in his life time. So he needs to use search engine to seek their desired information from the web. About 1.6 billion searches are performed in search engines every day. And a large number of time people have to search for multiple things one by one. As a result they have to type different search keyword multiple times. But it is a matter of stress. Often it is found that one peoples search pattern in a particular time matches other when other user falls in same type of situation. For example, before getting admission to a university people often search for different universities one by one. Instead of letting him search the entire site one by one if suggestion can be provided along with his initial search then it will make his life easier. To provide him appropriate suggestion with his search term, many approaches can be followed. If we provide him suggestion from non optimized table, It will be very time consuming and will turn out to be counterproductive as many user who doesn't really need the suggestions, need to wait few more microseconds so that the suggestion can be added. So ,we need to provide the suggestion but in such a way, that it can be beneficial to a large number of people who need the suggestion and is not much time killer so that those who doesn't need suggestion, doesn't suffer much for other parties benefit.

As a result we have proposed a model for improving the efficiency of this type of search where suggestions need to be shown along with result. By implementing the model the search area will be reduced and the user might rid of dual searches in some circumstances. Our model can handle large amount of data which are web addresses and optimize the web addresses according to their frequency by pattern mining technique and reduce the execution time of search.

There are three stages of our model following them we are actually achieving optimized data.

- Data collection
- Organizing data
- Mining data
- Optimization of data and its further use.

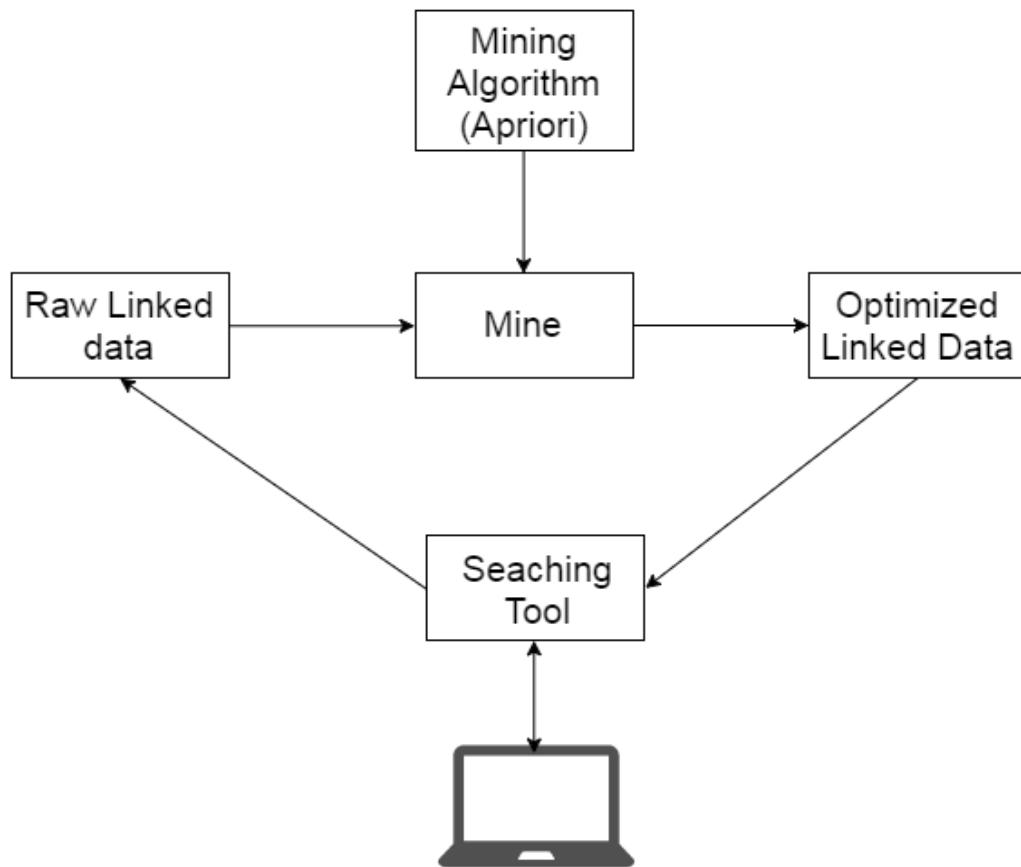


Figure 3.1: Proposed model for optimizing linked data

3.2 Data Collection:

The data is collected from different users. The search results of the users query are the web addresses. We are concern about the users visited websites. When the user clicks several links from search results we are actually take them as set of data and also taking the information of user (IP/username) because we are trying to identify the data sets individually.

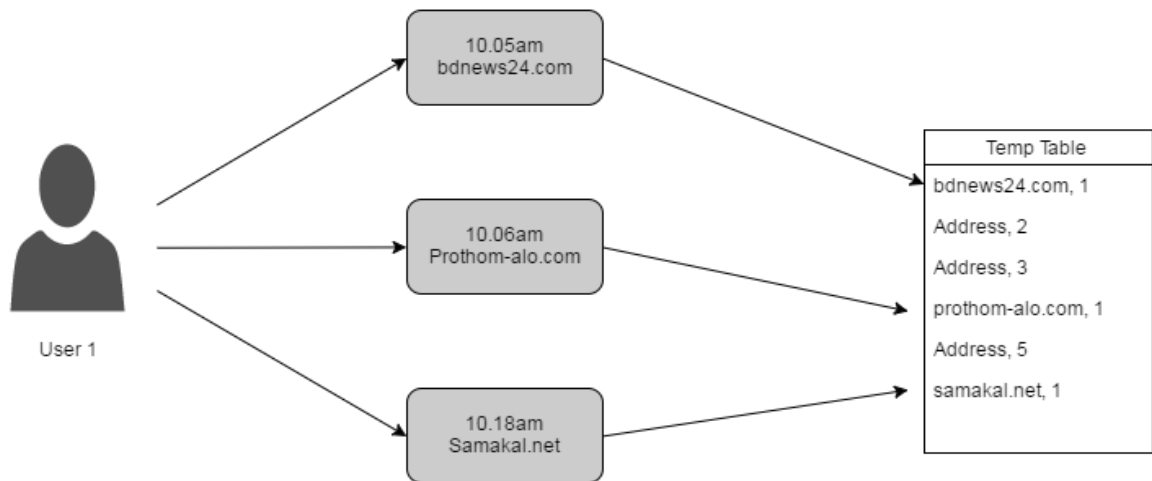


Figure 3.2: Data Collection from user.

3.3 Organizing Data:

As we have taken the data set from a particular user because we are assuming them the visited web addresses might have some relationship among them or their might have some objective behind searching the web addresses. So we have taken the set of web addresses as item set and organizing in this way it is helpful to implement it in our future algorithm.

Item set	Web link 1	Web link 2	Web link 3	Weblink4	Web link 5
1	www.ewubd.edu	www.du.ac.bd	www.ru.ac.bd	www.buet.ac.bd	www.sust.edu
2	www.mit.edu	www.stanford.org	www.stanford.edu	www.cuet.ac.bd	www.ruet.ac.bd
3	www.jnu.ac.bd	www.iitd.ac.in	www.ru.ac.bd	www.ewbd.edu	www.cuet.ac.bd

Table 3.1: Organizing data in item sets

3.4 Mining Data:

The data we have organized will be used for frequent pattern mining and from the different frequent pattern mining algorithm we have used Apriori Algorithm. The reason behind using Apriori is that, it is easier to implement among all other pattern mining algorithm and can easily be parallelized. Moreover it is most popular algorithm for mining frequent data set; as a result many people will be interested in further research of this project if it is submitted to code sharing sites as open source project. We have used Apriori algorithm with minor adjustment for our project as we are going to work with such combinations where order of occurrences doesn't matter. How we used Apriori in our project is shown below as a workflow diagram (code available in appendix):

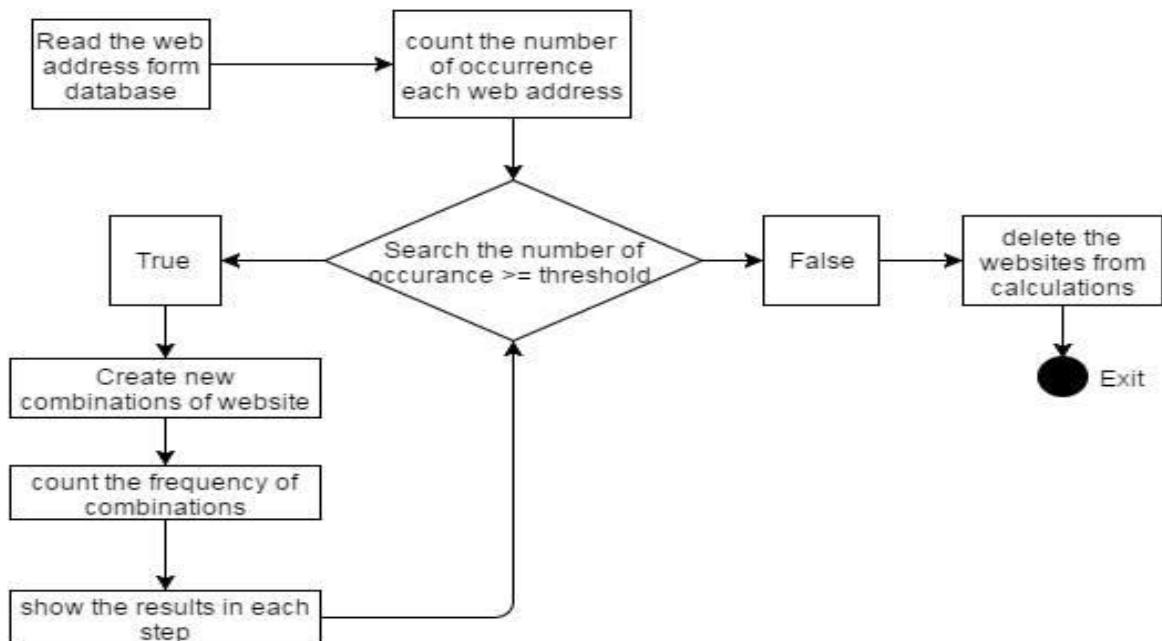


Figure 3.3: Implementation of Apriori algorithm for mining.

The Apriori algorithm can find the combination of websites for any number of degrees. But we are actually implementing 3rd degree optimization of the algorithm, because the data of 2nd degree is useful but the data is not reducing in good amount and might remain some redundancy which is not expected. If the 4th degree was chosen their might be huge chance of data loss according to the apriori technique. So 3rd degree is chosen in consideration of data is not redundant and accuracy of search results.

3.5 Optimized Data:

After the 3rd degree generation of website combinations, we have considered them as optimized linked data as the number of raw linked data is reduced by quite big margin. Now these data, which are retrieved by implementing Apriori, will be used for future search, which means, when user will search for any specific web address, at first the optimized data set will be inspected to show the user result. If found then the linked data set with the user’s desired address will be shown. In this way the search area is reduced and the user may find his other desired addresses in the linked data.

Web Address	Number of Occurrences	Web Address	Number of Occurrences	Web Address	Number of Occurrences
(www.ewu.edu)	5	(www.ewu.edu)	2	(www.ewu.edu)	3
(www.ewu.edu.bd)	3	(www.ewu.edu.bd)	3	(www.stanford.net)	
(www.stanford.net)	4	(www.stanford.net)	3	(www.stanford.edu)	
(www.stanford.edu)	3	(www.stanford.edu)	2	(www.ewu.edu)	
(www.mit.edu.bd)	2	(www.mit.edu.bd)	2	(www.stanford.net)	
(www.good.com)	2	(www.ewu.edu.bd)	2	(www.stanford.edu)	
		(www.good.com)	3		
		(www.stanford.net)			

Figure 3.4: Optimized data in different degree of Apriori.

Chapter 4

Implementation

Every project need to be tested and to test a project some type of tool is required to test the project on. Our project is based on user search, so it needs a searching tool to be tested. Most used searching tool in web is search engine. But most search engine owner keep their source code and functioning mechanism hidden for business reason. As a result we create our own searching tool to test our model on it. Implementation of our searching tool and how our model works on it will be discussed in this chapter. We can divide our implementation process into following parts :->

- Tools
- User Manual
- Client Side Implementation
- Server Side Implementation
- Outcome

4.1 Tools Description

4.1.1 Languages

We have mainly worked with languages of 3 different type:

1. For Programming-> PHP, JavaScript, cURL .
2. Database Management->MySQLI .
3. Visualization->HTML,CSS,Bootstrap .

Programming

We have used PHP as our major programming language on Xampp platform which transform our pc into a local server. PHP is a widely-used open source general-purpose scripting language (scripting language is a subset of programming language that is used to produce scripts, which are sets of instructions that automate tasks that would otherwise be performed manually by a human.) that is especially suited for web development. We have also used Javascript and CURL for few small units of our project.

Why Scripting Language Over Programming Language?

- Our Searching tool needs frequent interaction with internet and using programming language for that means extra line of codes need to be typed.
- PHP is server-aligned and most programming languages are not.
- Maintaining a PHP+Apache server is far simpler than maintaining a TomCat or Glassfish server. [8][12]

Why PHP?

- PHP is most popular server-sided language.
- PHP results in faster site loading speeds than any other scripting language.
- It is Open Source.
- PHP is flexible for database connectivity. It can connect to several databases the most commonly used is the MySQL.
- PHP has very good online documentation with a good framework of functions in place.
- PHP Accepts Platform Diversity.[6][8]

JavaScript and cURL

JavaScript is known as programming language of the Web. Main difference between JavaScript and PHP is that PHP is server side language where JavaScript is Client Side Language. We have used JavaScript for passing few data to a php page so that user activity can be stored for mining in future. And to grab important information out of soon to be added websites for storing to database. [10]

cURL, also known as Client URL is a tool to transfer data from or to a server, using one of the supported protocols (DICT, FILE, FTP, FTPS, GOPHER, HTTP, HTTPS, IMAP, IMAPS, LDAP, LDAPS, POP3, POP3S, RTMP, RTSP, SCP, SFTP, SMB, SMBS, SMTP, SMTPS, TELNET and TFTP). The command is designed to work without user interaction. In our project, rather than using command line cURL, we are using libcurl, a library created by Daniel Stenberg which is supported by newer version of PHP (4.0.2+) which creates a bridge to use command line cURL from the php language. We have used libcurl to allow us to use a variety of web resources from within our PHP script. [6][10]

4.1.2 Database Management

We have used MySQLI to manage our database. It stands for Improved Structured Query Language. It is the improved version of most popular Open Source SQL database management system, MySQL.

Why MySQLI ?

- Performs better than other Query Language when used with newer version of PHP.
- Support for prepared statements which was not available in MySQL.
- Provide Object-oriented interface.
- Enhanced debugging capabilities.
- Embedded server support. [6][8]

4.1.3 Visualization

As our project runs on a browser .So, for visualization we used Hypertext Markup Language (HTML) to create basic structure of the visible pages we needed. We also used cascading style sheet (CSS) and Bootstrap framework for designing the page. Bootstrap framework was also used to make the page responsive according to the device size.[13]

What we used to make the Searching Tool

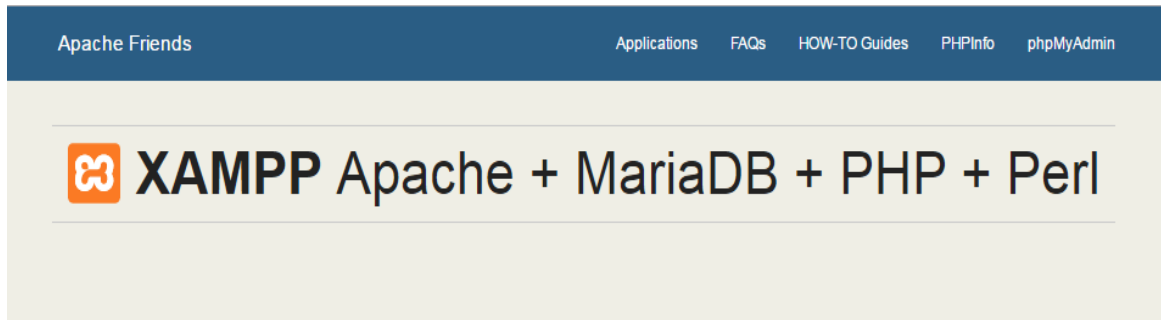
To write programs we have used simple text editor ‘Sublime Text’.



```
searchn.php x
a</td><td>",$row[1], "</td></tr>"; //title*/
235 echo "<tr class='success'><td><a onclick='storelink(\".$row[0].\", \"$.get_client_ip().\", \"$.SESSION[\"
    usernamesession\".\".\")\">",$row[0]. "</a></td><td>",$row[1], "</td></tr>"; //title
236 $flag=$row[0];
237
238 $fnresult0=mysqli_query($con,"SELECT id from website where webadd=".$row[0].";") or die("Final result
    Wrong Query00");
239 $fnrow00 = mysqli_fetch_row($fnresult00);
240 //echo $fnrow00[0];
241 //$fnresult0=mysqli_query($con,"SELECT distinct linkedata1 , linkedata2, linkedata3 FROM fainalresult WHERE
    linkedata1=".$fnrow00[0]. " or linkedata2=".$fnrow00[0]. " or linkedata3=".$fnrow00[0].";") or die("Final
    result Wrong Query0");
242 $fnresult0=mysqli_query($con,"SELECT linkedata1 from fainalresult WHERE linkedata1=".$fnrow00[0]. " or
    linkedata2=".$fnrow00[0]. " or linkedata3=".$fnrow00[0]. " UNION SELECT linkedata2 from fainalresult
    WHERE linkedata1=".$fnrow00[0]. " or linkedata2=".$fnrow00[0]. " or linkedata3=".$fnrow00[0]. " UNION
    SELECT linkedata3 from fainalresult WHERE linkedata1=".$fnrow00[0]. " or linkedata2=".$fnrow00[0]. " or
    linkedata3=".$fnrow00[0].";") or die(mysqli_error($con));
243
```

Figure 4.1: Sublime Text Editor Coding Window

And to test our project we used Apache Local Server, PHPmyadmin tool both of which came as a package with free and open source cross-platform web server solution software called XAMPP.



Welcome to XAMPP for Windows 7.0.6

You have successfully installed XAMPP on this system! Now you can start using Apache, MariaDB, PHP and other components. You can find more info in the FAQs section or check the HOW-TO Guides for getting started with PHP applications.

Start the XAMPP Control Panel to check the server status.

Figure 4.2: Xampp Server

4.1.4 Support for Running Our Program

As we are using scripting language, so no special integrated development environment needed. Only a simple browser on Client Side is good enough to run our Searching tool, But PHP and MySQL must be installed on server side.

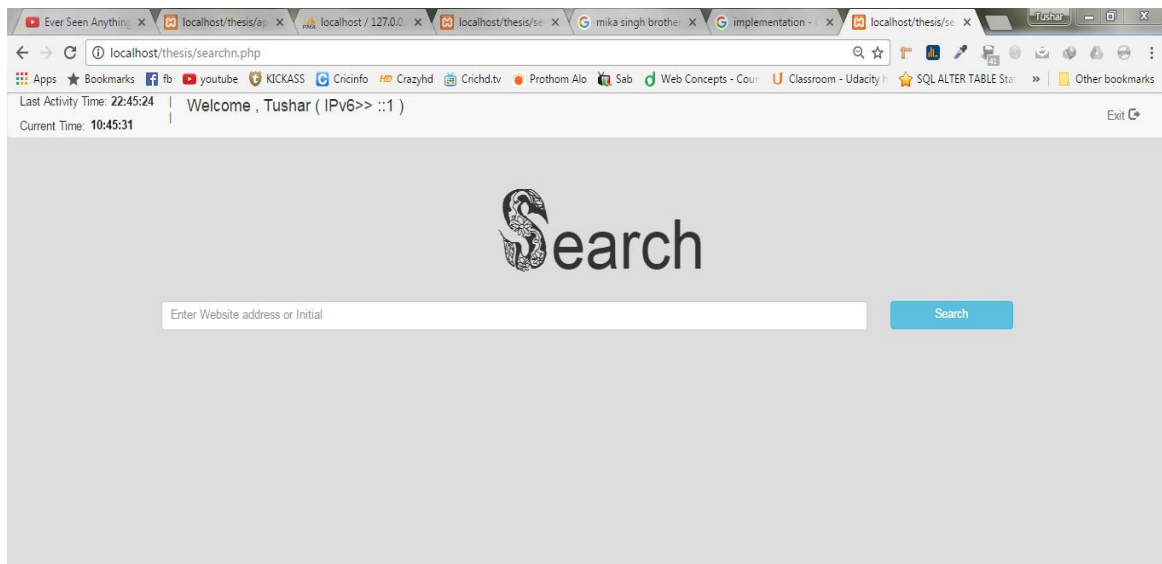


Figure 4.3: Visualization of program in a browser.

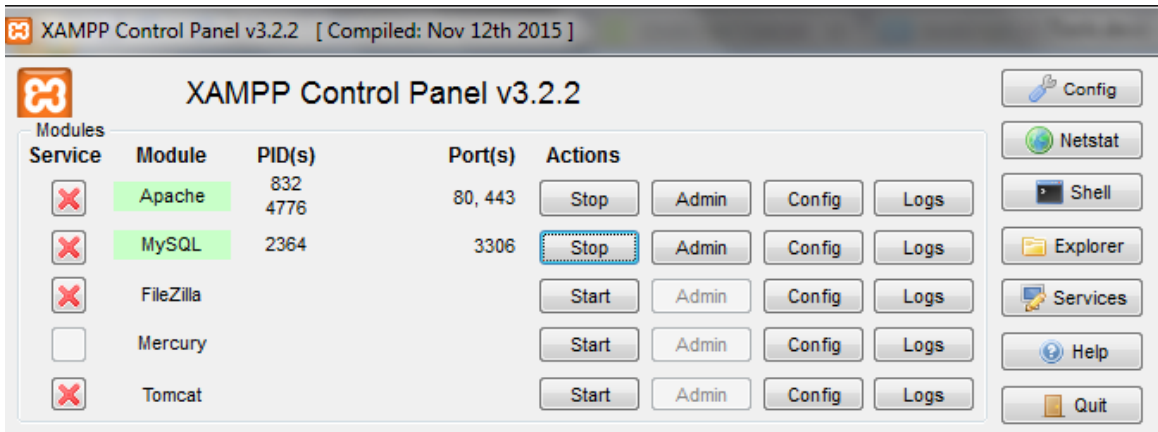


Figure 4.4: Active Server and Database

4.2 User Manual

User can use the tool like any other search engine. Only thing he has to do is login using any username. No password required as no private information sharing or person specific benefit is available. He just has to search for desired website. He can search by web address, title and even initial name. Once result appears, he can click on his desired link and enjoy the benefit. Rest of the work will be taken care of by the tool itself.

4.2.1 Login Procedure:

When a user wants to perform any search, first he/she has to go to login page. The login page is designed like the following picture->

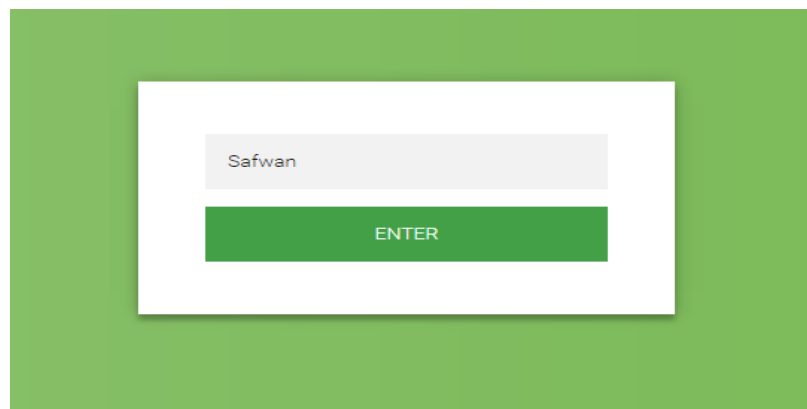


Figure 4.5: Login menu.

Only 1 text box is visible in this page and it asks for username. user have to put his username here. If any user is concerned about his security he can enter using a dummy name too even though real name is preferred.

4.2.2 Searching Techniques:

User doesn't need to be concerned about how the search is going on or what type of structure he has to follow in this tool. It will be specially done by the tool itself. A person just has to enter a search term in textbox and press the search button on screen or the enter button on keyboard. After that if result is available in database then it will show the result quickly but if not then user may need to wait for few seconds, that's all. Search result will appear once the searching is done.

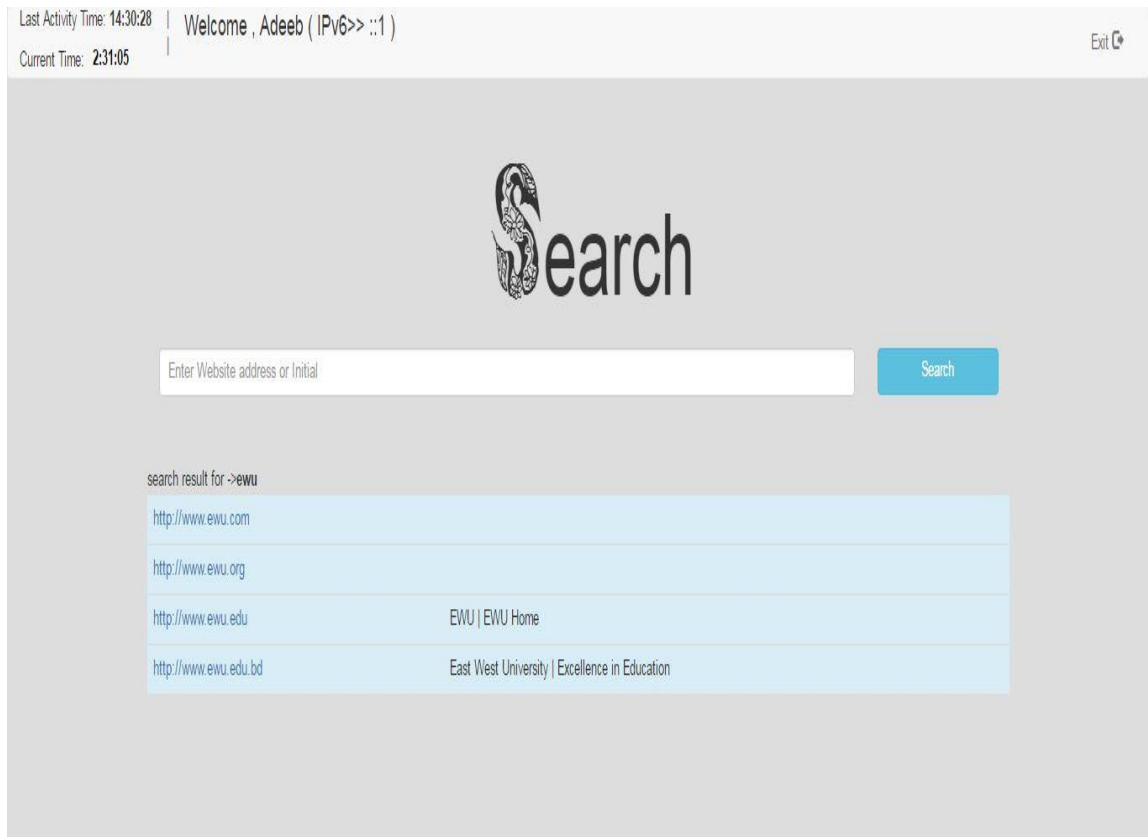


Figure 4.6: Performing Search

4.3 Client Side

4.3.1 Performable Search Type

When a user gives a query of a website he/she might search by the name of university or initial of the university or the full web address. So we are considered these facts to show the web addresses to the user which he is searching for. All the condition of the searches will be run to find out the website addresses. There is a website table in our database where we are storing the university names along with their id (which is auto incremented) and their title.

Title Search:

When a user gives a title of the university in the input field and searches for it, the sql query will be run which searches in the website table and looks for the match of the website titles of the universities and finds out the web addresses which titles are matched with the users given title. For example if any one searches for “East West University” then the web address shown to the user is ewubd.edu. Because, It has a title containing ‘East West University’ inside the website table.

Exact Match:

When a user gives any web address for searching the website there is a SQL which searches in the website table to find out the website address which matches with the user’s given web address. For example, if a user gives ewubd.edu then the tool show the web address from the website table which address in ewubd.edu.

String Match:

If the user does not know the actual web address, he might search by the initial of the web address, then tool will try to find out the web address which contains the initials in the full web address. Then it will Show the web address which contains the initial which matches the users input. For example the user searches for only “ewubd” then the ewubd.edu will be shown to the user because this web address contains the initial “ewubd”.

Online Search:

If all the above types fails to generate web addresses then this procedure will run to show web addresses. Firstly different extensions will be added to create web addresses and check if the created web site is valid or not. Then it will show the valid web addresses to the user which are actually exists in the online with their title.

Here the procedure is discussed briefly,

- **Adding extensions:**

We have several web extensions in our database which we will add in the rear of the given initial by the user and we will generate several web addresses to figure out the desired one. The extension we are adding are .org, .com, .gov , .bd , .edu etc. then these extensions will create several web addresses from one initial.

- **Check validity:**

get_headers() function:

For checking the validity of the website, using get_headers() function is a solution to check the validity of a web sites. get_headers() function fetches all the headers sent by the server in response to a HTTP request. It returns an array where different index of the array has different values which are shown below.

```

Array
(
    [0] => HTTP/1.1 200 OK
    [1] => Date: Sat, 29 May 2004 12:28:13 GMT
    [2] => Server: Apache/1.3.27 (Unix) (Red-Hat/Linux)
    [3] => Last-Modified: Wed, 08 Jan 2003 23:11:55 GMT
    [4] => ETag: "3f80f-1b6-3e1cb03b"
    [5] => Accept-Ranges: bytes
    [6] => Content-Length: 438
    [7] => Connection: close
    [8] => Content-Type: text/html
)

```

Figure 4.7: The Array index and their values

here the index [0] is our main consideration. If the web address is actually exists, There will be “true” result in the [0] index of the array.

The client device will sending an HTTP request to the server of the web addresses. If the server responses with true result, we will considers it as a valid web address.

These HTTP request is continuously send to all the servers form the user’s device and waiting for the response for a certain period of time. After the certain time the tool shows the web addresses which are valid. [6]

- **cURL:**

Using the `get_headers()` function is a fair way to check the validity of the websites but it has a problem of time consumption. The user have to wait until the timeout of the `get_headers()` function if a particular site doesn’t exist and cant return the header file. So in replacement, we have used cURL.

curl is a tool to transfer data from or to a server, using one of the supported protocols (DICT, FILE, FTP, FTPS, GOPHER, HTTP, HTTPS, IMAP, IMAPS, LDAP, LDAPS, POP3, POP3S, RTMP, RTSP, SCP, SFTP, SMB,

SMBS, SMTP, SMTPS, TELNET and TFTP). The command is designed to work without user interaction.

Even though curl is unusable inside php but PHP supports libcurl, a library created by Daniel Stenberg, that allows us to connect and communicate to many different types of servers with many different types of protocols. libcurl currently supports the http, https, ftp, gopher, telnet, dict, file, and ldap protocols. libcurl also supports HTTPS certificates, HTTP POST, HTTP PUT, FTP uploading (this can also be done with PHP's ftp extension), HTTP form based upload, +proxies, cookies, and user+password authentication.

These functions have been added since PHP 4.0.2. [6][10]

How cURL Improves from PHP's get_header():

The curl_setopt() function grabs the header of the web addresses from the servers. If the website exists the curl_setopts() function can grab the header of the website if it can't , the website doesn't exists. We have used that methodology to validate a website. This method reduces time consumption because there is no continuous packet sending in the cURL. So it reduces the time of validity checking. Approximately using cURL method responses 3 times faster than the get_headers() function method. So we are these was the main reason for implementing cURL instead of get_header() method.

▪ Multithreading in cURL:

The Curl can be used for sending multi requests simultaneously. The curl_multi_*(*) functions can handle multiple response from the servers at a time. So using these multithreading technique the validity checking of the websites will be reduced .For our project it was not possible to implement because of the limitation ofxampp local server. But in actual server it will work just fine. [6][8]

4.3.2 Input Acquisition

People use search engine for various reason. Our searching tool works like a search engine. People can search for various websites based on their need. And our Searching tool provides them their desired result along with recommended website for a particular site based on how frequently other people searched for that website. To do that we need to mine older data and generate desired frequently searched combination. At first, we need raw data to apply mining on. So, input data generated by users, need to be acquired for processing.

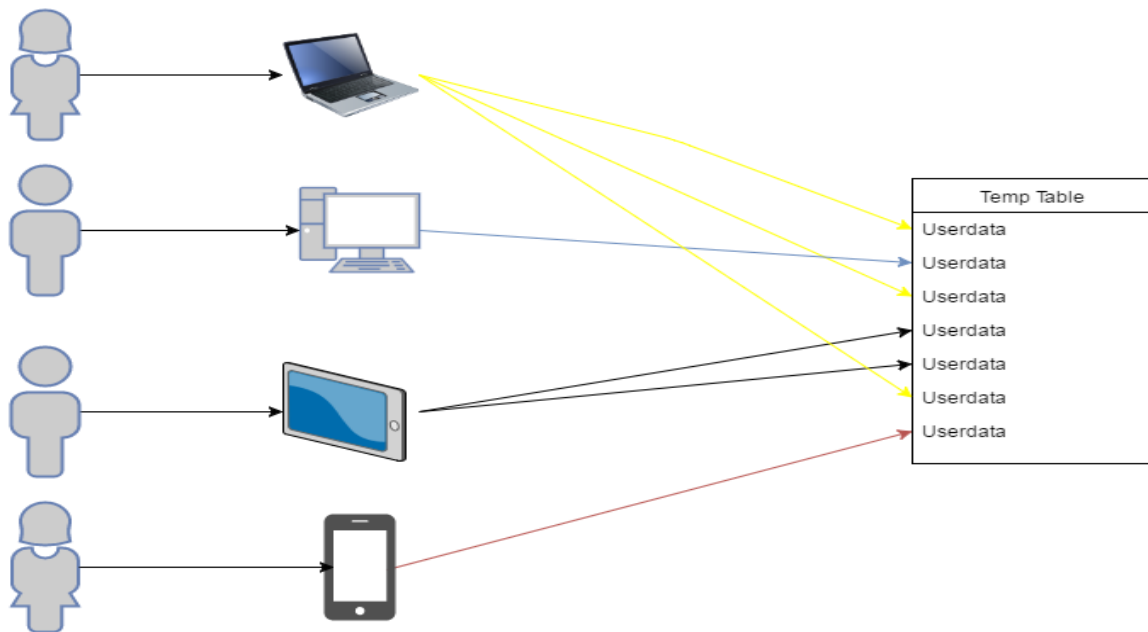


Figure 4.8: 'Input Acquisition'

While a person is logged in and searching for different websites a table at database will collect those websites along with user identity temporarily. And every new data found from those result is stored in website table where each website gets a unique website id.

			id	webadd	univ_name
<input type="checkbox"/>	Edit	Copy	Delete	1	www.ewu.com
<input type="checkbox"/>	Edit	Copy	Delete	2	www.ewu.org
<input type="checkbox"/>	Edit	Copy	Delete	3	www.ewu.edu EWU EWU Home
<input type="checkbox"/>	Edit	Copy	Delete	4	www.ewu.edu.bd East West University Excellence in Education
<input type="checkbox"/>	Edit	Copy	Delete	5	www.stanford.net
<input type="checkbox"/>	Edit	Copy	Delete	6	www.stanford.edu Stanford University
<input type="checkbox"/>	Edit	Copy	Delete	7	www.mit.edu MIT - Massachusetts Institute of Technology
<input type="checkbox"/>	Edit	Copy	Delete	8	www.mit.edu.bd (MIIT)
<input type="checkbox"/>	Edit	Copy	Delete	9	www.good.com BlackBerry Enterprise Mobility, BBM, Smartphones &...
<input type="checkbox"/>	Edit	Copy	Delete	10	www.good.org Homepage & middot; Good Samaritan United Methodist ...
<input type="checkbox"/>	Edit	Copy	Delete	11	www.good.net

Figure 4.9: 'website' table

4.3.3 Differentiating Between Users

User identification is required to distinguish between users who are using the searching tool at the same time from different communication devices. Here we have used both username and IPv6 of the device for identification. We excluded any need of password to reduce complexity and as no special functionality available for specific user. Collecting user IP is a little hard due to the fact of using wireless network and shared IP system or a proxy server. But we still managed to collect IP using the function `get_client_ip()`. Here is the pseudo code of IP collecting function.:-

```
Function get_client_ip() {
    Detect type of Ip
    if shared ip
        return public_ip;
    else if proxy server
        return ip_fowarder;
    else
        Return remote_address;
```

```
}// full code available on Appendix section
```

uname	uip	id
Tushar	::1	3
Tushar	::1	4
Tushar	::1	2

Figure 4.10: 'temptable' table

The picture above shows the table in which a person who is accessing website using IPv6 ::1 (localhost) and all 3 websites he searched for in a single session. These data is not yet processed or ready for processing. When a user log him/her out or a predefined amount of time has passed since he last searched a website (2 hour in our case) only then these data will be ready for processing.

4.4 Server Side

4.4.1 Session:

When a user logs in a Session is created for the current user which we are coded in PHP. Automatically a session is created by the user name and also a time section in the web page which shows the current time and last activity time. When a user logs out the session is destroyed and the websites which are clicked, considered as transaction/set and it is used for as future Apriori calculation. The session destroys automatically when the page remains idle for a certain period of time. It can be changed as requirement of developer. Session is important for this project because we need several set of websites.

4.4.2 Collecting Raw Linked data

After a user is done with his particular session, all his data from temporary table will be represented as a set of websites in a table called transaction where this set will be represented in a 2 column table where 1st column is an auto generated id of that particular combination and the 2nd column is id of a particular website. For 'N' number of search in such a set will generate 'N' number of row in transaction

table and those data are deleted from temptable. If a user only search for 1 website and session ends then that data is not required for mining as it doesn't create any combination with other website on that instance, so that data from temptable is deleted without adding to transaction table.

tid	webid
1	3
1	6
1	5
2	3
2	6
2	5
3	9
3	8
3	5
3	4

Figure 4.11: 'transaction' table

This figure above shows the representation of combination set in database table 'transaction' where tid is auto generated id of a combination and webid is one of member website's id on that combination. Following code is responsible for doing these job before a session_destroy() function is called:

```

session_destroy() {
    Collect username
    Collect IP
    Collect all data for that IP and username
    If amount of data less more or equal degree of
    apriori
        Generate an id for combination.
        While any data left on temporary table
            add(combinationid, website)to transaction
            table
        Delete all data used found for that username and
        IP from temporary table
    }

```

```

    End session
}
//Full code on Appendix part of this paper

```

4.4.3 Effect of Mining Algorithm

The data on transaction table is the data where mining algorithm is applied. After mining algorithm is applied we see its effect on output, when a user is searching after the mining process. We applied the apriori algorithm in our tool as it is well known and most commonly used frequent data analyzing algorithm. The workflow diagram of Apriori algorithm given below->

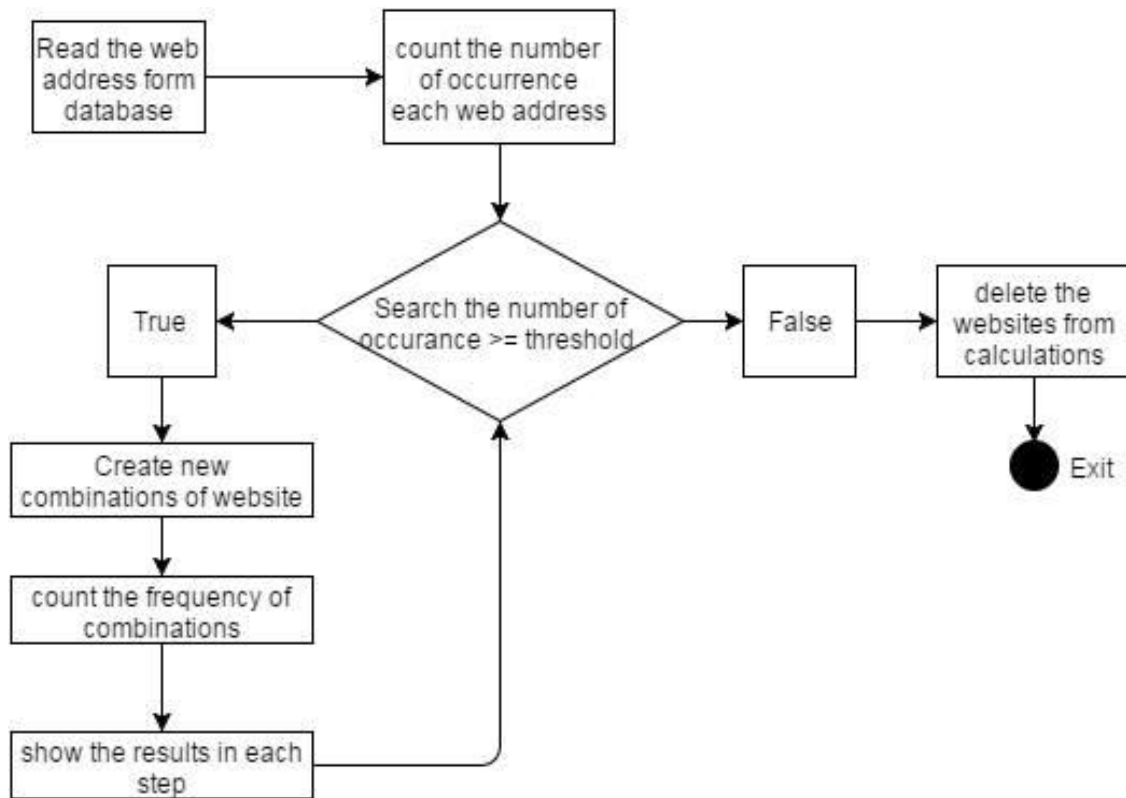


Figure 4.12: Workflow of Apriori algorithm

(Full code added at appendix part of this paper)

Here are the changes of result after each degree of apriori algorithm applied →

Effect of 1st degree Mining Algorithm

1st degree of apriori algorithm checks occurrence of each website on ‘website’ table of our database. If the occurrence is more than certain threshold (2 in our example) only then it is added to generate candidate set for next degree of mining, if not those websites are excluded for further calculation. After applying 1st degree of apriori algorithm on our data it shows us which websites are frequently visited and which websites are rarely visited. Rarely visited websites are less important as most people are interested in sites with more popularity. So if we apply only 1 degree of apriori we can improve suggestion by excluding less visited site. But this is not good enough as we don’t know when to suggest what.



1

Figure 4.13: 1st degree apriori applied on ‘transaction’ table

Effect of 2nd degree Mining Algorithm

2nd degree of apriori algorithm checks occurrence of each 2 website combination from 'website' table of our database. These combination need to be generated. We know from 'N' members in candidate set we can get $(n-1)+(n-2)+\dots+1=(n*(n-1))/2$ number of possible combinations. If the occurrence is more than certain threshold (2 in our example) only then it is added to generate candidate set for next degree of mining, if not those combination are excluded from further calculation. After applying 2nd degree of apriori algorithm on our data it shows us which 2 website combinations are frequently visited and which 2 website combinations are rarely visited. If we apply 2nd degree of apriori we can improve suggestion by excluding less visited websites and also less visited combinations. It is good enough to provide suggestion when a website is searched based on its existence on optimized table 'finalresult' by searching other half of its combination. Even though this is great improvement from 1st degree apriori but there are still room for improvement as too much suggestion may populate search result area.



Figure 4.14: 2nd degree apriori applied on 'transaction' table

Effect of 3rd degree Mining Algorithm

3rd degree of apriori algorithm checks occurrence of each 3 website combination from 'website' table of our database. These combinations also need to be generated. Here to generate combination we do join combination from 2nd degree candidate set. Still, from 'N' members of 2nd degree combination in candidate set we can get $(n-1)+(n-2)+\dots+1=(n*(n-1))/2$ number of possible 3 website combinations. If the occurrence is more than certain threshold (2 in our example) only then it is added to generate candidate set for next degree of mining, if not those combination are excluded from further calculation. After applying 3rd degree of apriori algorithm on our data it shows us which 3 website combinations are frequently visited and which 3 website combinations are rarely visited. If we apply 3rd degree of apriori we can improve suggestion by excluding less visited websites and also less visited combinations like 2nd degree apriori. Moreover it reduces the populated suggestion area by suggesting more accurate result on 2nd search. If a person search for a site and again search for another site and both sites are in a 3website combination after 3rd degree apriori applied on raw linked data then that user is probably going to be searching for the 3rd site from that combination site as well. So on that time we can suggest them the 3rd website from the combination and reduce the number of suggestion provided to user. Adding more and more degree to this algorithm minimize the number of suggestion provided to user but gradually decreases the accuracy so in our tool we implemented up to 3rd degree to get best possible result.



Figure 4.15: 3rd degree apriori applied on 'transaction' table

4.5 Outcome

Final result after performing Apriori algorithm to mine those 3rd degree linked data is stored in a table called 'finalresult'. This table has a direct one directional connection from server to user. This table is exclusively used to serve user with optimized result by not letting user access overloaded transaction table to provide search suggestions.

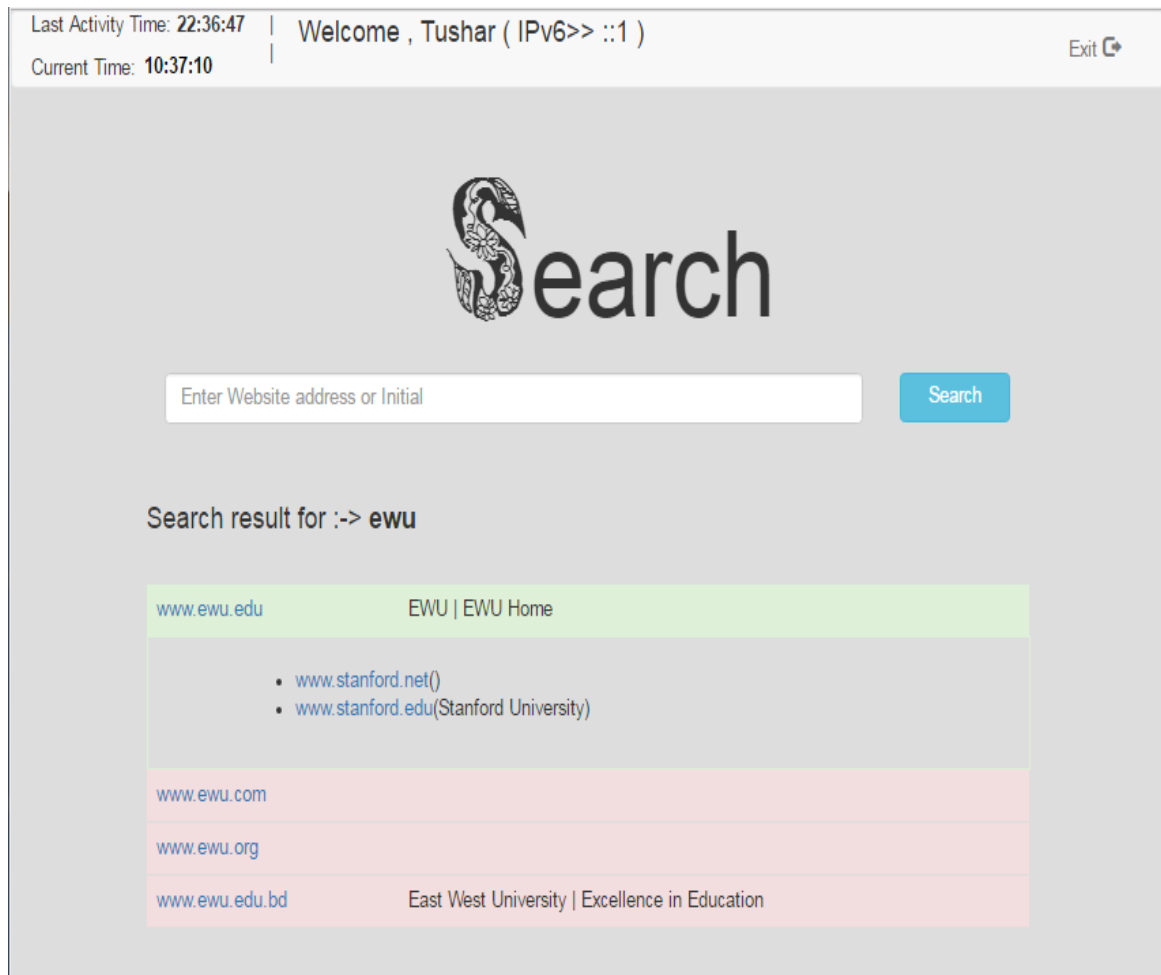


Figure 4.16: User getting benefit of optimization

If we would like to do the same without using frequent item set mining algorithm this process of providing suggestion would take long time as the program need to traverse all field in transaction table and also require a very complex query to find desired link between websites. But By using frequent item set mining algorithm the number of field need to be traversed in number of degree of algorithm we used*number of combination found. Even though multiplication sign exist in 2nd situation but according to real life situation number of linked combination are so much less than total available raw combinations as a result, making it many times more optimized than before. Even if raw data's were collected by extreme filtering and categorization it will still remain very efficient while providing suggestion, and thus minimizing optimizing overall search result generation.

Chapter 5

Result Analysis

As we used apriori algorithm for our project so we get quite accurate result, but the process of applying apriori very time consuming to generate all possible combinations as well. But still, it doesn't raise any question about the effectiveness of our model as the algorithm doesn't need to run on real time. The algorithm can be applied on stored data on a certain time of day/week or even month. And user can get benefited from it, once it is applied.

Before our model was applied a search engine has to search all the cell to find the desired website using their default searching mechanism(mysql's default, in our case) and to find its linked website it use to need searching all the combination sets and all its elements ,which was very time consuming. But now after the model was applied, the cell need to be traversed has reduced by far.

After taking data from 200 users, it was found that 103 people searched more than 2 website after they logged in a particular session. As 97 people searched for 2 or less number of websites in a single session, so there data is excluded from calculation and deleted from that table of database as it is impossible to generate 3 combination sites from a set which consist of 2 or less member. From eligible 103 user's search history, 340 input data (raw linked data) was collected from those who searched more than 1 website in a particular session. After that, the mining algorithm was applied and the result was stored in another separate table so that system doesn't need to go though the large raw code to answer a call from user. The end result is quite satisfactory. Here is the comparison between necessity of cell traversal before and after applying proposed model with apriori→

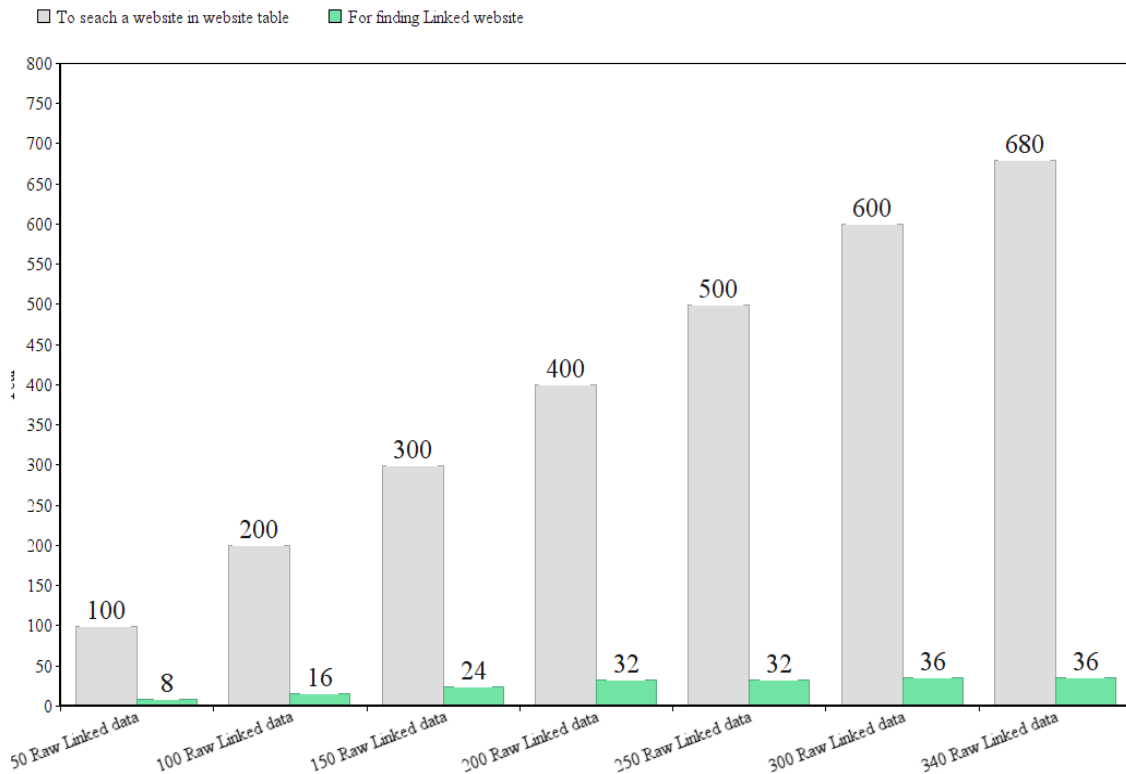


Figure 5.1: Comparison between optimized and un-optimized data

Here in y axis,

Grey colored bar=Number of cell traversal before optimization.

Pastel green colored bar=Number of cell traversal after optimization.

And in x axis,

Number of raw linked data for experiment is shown as grey, green couple where the color represents

Even though this is not the perfect scenario every time, as the number of input data is given by user and thus the optimized data also clearly depends on that. But in most case, the result is quite similar and even in worst case the number of cells in optimized result is only a few more than normal case

After the comparison between read lengths for suggestion generation, now let's check the overall outcome when a user searches for a site how many cell need to be traversed in total to answer his call? The result depends on how many websites

are on the database. In our case we tested the system keeping the number of overall website 400 (constant). The overall cell traversal required is shown on the table →

Cell's need to be Traversed			
Number of Input data	Before Optimization	After Optimization	
		If Available in Optimized data	If Not Available in Optimized data
100	$400+200=600$	16	$600+16=616$
200	$400+400=800$	24	$800+24=824$
300	$400+600=1000$	32	$1000+32=1032$
340	$400+680=1080$	36	$1080+36=1116$

Table 5.1: Cell traversal data computation

So, for our particular scenario, after optimization, we get huge improvement in terms of number of cells need to be traversed. But in worst case when a search is performed and the desired data is not in optimized data table, we also see very small amount of performance decrease, which is a limitation of this approach. But as all of the optimized data are more frequent so it is expected that most people will search for those data which are already in the optimized data table. So in most case we will get huge performance increase and only in few cases we will see a little performance drop. So, overall our proposed model will be very efficient compared to current searching system and especially for linked data analysis the impact of our proposed model is huge.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

Now the in modern world, the rapid change of technology is a natural trend. There may be several platforms which may exist; it may go high in the future. In such rapid changing situation of technology, it is very uncertain to tell which platform will dominate the market. Computer science is the study of the theory, experimentation, and engineering that form the basis for the design and use of computers. More succinct definition of computer science is the study of automating algorithmic processes that scale. A computer scientist specializes in the theory of computation and the design of computational systems. So, future researchers may invent new technology or ideas to improve the optimization techniques of frequent data. In our perspective the search optimization is beneficial for giving suggestions and showing the relation among the data.

As the modern technology is updating continuously it is common issue that new technology of searching technique might be induced in the reduction of searching optimization. It is hard to tell that how effectively frequent data mining will help to revolutionize the industry. Current search engines have reached such a place where cost, reliability and throughput are very important factor. And with time more and more people will be involved in web activity that, linked data mining for search optimization can be significant part of a search engine.

6.2 Future Work

Lots of elements in this project leave scope for further development. Few general areas will be highlighted here →

- Assessment of other mining algorithm to check if that is better suited algorithm for the project.
- Adjusting the system for big data in distributed computing platform.
- Adapting searching tool according to statistics.
- Replacing current data grabbing logic with smarter Information grabber.
- Using multithreading for quicker Information grabbing.
- Finding better searching technique according to need of this project.
- Analyzing the impact of using other type of databases like graph database, multi valued or object oriented database rather than relational database.

Bibliography

1. Dr. Nagaratna Hegde, “Web Search Result Optimization using Association Rule Mining Algorithm” Published on: IJSET - International Journal of Innovative Science, Engineering & Technology, December 2015.
2. Rashmi Jha, “Association Rules Mining for Business Intelligence “Published on: International Journal of Scientific and Research Publications, May 2014.
3. W. Bruce Croft, Donald Metzler,Trevor Strohman, “Information Retrieval in Practice” , 2015
4. Dr. J.S.Sodhi, Amita Joshi“Target Advertising via Association Rule Mining”, Published on: International Journal of Advance Research in Computer Science and Management Studies, May 2014.
5. Sudhkar Singh, Apriori Based Algorithms on Map Reduce Framework Published on: International Conference on Communication and Computing, ICC 2014.
6. URL: <http://php.net/manual/en/> [Last accessed: 20/11/2016]
7. URL: <https://www.researchgate.net/> [Last accessed :2/12/2016]
8. Website: <http://stackoverflow.com> [Last accessed:25/11/2016]
9. URL: <https://www.youtube.com/user/nptelhrd> [Last accessed :20/12/2016]
10. URL:<https://curl.haxx.se/docs/manpage.html> [Last accessed :28/12/2016]
11. Website: <https://www.wikipedia.org>
12. Website: <https://youtube.com>
13. Website: <http://www.w3schools.com/>
14. Book: ‘Data Mining - Concepts and Techniques (3rd Ed)’ by Jiawei Han, Micheline Kamber, Jian Pei

APPENDIX

A1. Home Page

```
<?php
    session_start();
    if(isset($_SESSION["username"]))
    {
        header("Location: searchn.php");// if logged in, not
allowed here.
    }
?>
<!DOCTYPE html>
<html >
<head>
<meta charset="UTF-8">
<title>Home</title>
<link rel="stylesheet" href="css/style.css">
</head>
<body>
<div class="login-page">
<div class="form">
<form class="login-form" method="post"
action="searchn.php">
<input type="text" name="un" placeholder="username">
<button>Enter</button>
```

```
        </form>
</div>
</div>
</body>
</html>
```

A2. Collecting User IP

```
function get_client_ip() {
    $ipaddress = 'nothing';
    //if visitor using router(shared internet) than return
    real ip
    if (getenv('HTTP_CLIENT_IP'))
        $ipaddress = getenv('HTTP_CLIENT_IP');
    //if proxy server exist in between connection
    else if(getenv('HTTP_X_FORWARDED_FOR'))
        $ipaddress = getenv('HTTP_X_FORWARDED_FOR');
    else if(getenv('HTTP_X_FORWARDED'))
        $ipaddress = getenv('HTTP_X_FORWARDED');
    else if(getenv('HTTP_FORWARDED_FOR'))
        $ipaddress = getenv('HTTP_FORWARDED_FOR');
    else if(getenv('HTTP_FORWARDED'))
        $ipaddress = getenv('HTTP_FORWARDED');
    //regular ip address collection
    else if(getenv('REMOTE_ADDR'))
```

```

        $ipaddress = getenv('REMOTE_ADDR');
else
        $ipaddress = 'UNKNOWN';
return $ipaddress;
}

```

A3. Searching Mechanisms

```

$result0=mysqli_query($con,"SELECT webadd,univ_name,id
FROM website WHERE univ_name LIKE '%".$initial."%")or
die("WrongQuery0");

```

```

$result1=mysqli_query($con,"SELECT webadd,univ_name,id FROM
website WHERE webadd='".$initial."'")or die("Wrong
Query1");

```

```

$result2=mysqli_query($con,"SELECT webadd,univ_name,id FROM
website WHERE webadd LIKE '%".$partofurl[0]."%")or
die("Wrong Query2");

```

```

$result3=mysqli_query($con,"SELECT webadd,univ_name,id FROM
website WHERE webadd LIKE '%".$initial."%")or die("Wrong
Query3");

```

```

$flag=99999999;

```

```

while ($row0 = mysqli_fetch_row($result0)) {
    echo "<tr
class='success'><td><aonclick=\"storelink('".$row0[0].
','.$_SERVER['REMOTE_ADDR'].','.$_SESSION['usernameession']
')."\">
>".$row0[0]."</a></td><td>",$row0[1],"</td></tr>";

```

```

$flag=$row0[0];

```

```

}

while ($row1 = mysqli_fetch_row($result1)) {

echo"<tr class='warning'><td><a
onclick=\"storelink('".$row1[0]."', '".$get_client_ip()."', '$_SESSION["username"]')\"
>".$row1[0]."</a></td><td>",$row1[1],"</td></tr>"; //exact
match

$flag=$row1[0];

$fnresult02=mysqli_query($con,"SELECT id from website where
webadd='".$row1[0]."'");or die("Final result Wrong
Query00");

$fnrow01 = mysqli_fetch_row($fnresult02);

$fnresult1=mysqli_query($con,"SELECT linkedata1 from
fainalresult WHERE linkedata1=".$fnrow01[0]." or
linkedata2=".$fnrow01[0]." or linkedata3=".$fnrow01[0]."
UNION SELECT linkedata2 from fainalresult WHERE
linkedata1=".$fnrow01[0]." or linkedata2=".$fnrow01[0]." or
linkedata3=".$fnrow01[0]." UNION SELECT linkedata3 from
fainalresult WHERE linkedata1=".$fnrow01[0]." or
linkedata2=".$fnrow01[0]." or
linkedata3=".$fnrow01[0].";")or die(mysqli_error($con));

if (mysqli_num_rows($fnresult1)>0)

{

echo "<tr><td colspan='3' style='padding:20px 80px;
border:2px solid #DFF0D8;'><ul>";

while ($fnrow1 = mysqli_fetch_row($fnresult1)){
    if($fnrow1[0]!=$fnrow01[0]){

        $fnresult03=mysqli_query($con,"SELECT webadd,univ_name
from website where id='".$fnrow1[0]."'");or die("Final
result Wrong Query01");

        $fnrow1 = mysqli_fetch_row($fnresult03);

        echo "<li><a
href='#'>".$fnrow1[0]."</a> ( ".$fnrow1[1].")</li>";}}

```

```

    }

while ($row3 = mysqli_fetch_row($result3)) {
if($flag!=$row3[0])//to exclude exact match to show again
{
    echo "<tr class='danger'><td><a
    onclick=\"storelink('".$row3[0]."', '".get_client_ip().
    "','".$_SESSION["username"]."')\"
    >".$row3[0]."</a></td><td>",$row3[1], "</td></tr>";}
}

if(mysqli_num_rows($result3)==0)
{
while ($row2 = mysqli_fetch_row($result2)) {
    if($flag!=$row2[0])//to exclude exact match to show
again
    {
        echo "<tr class='info'><td><a
        onclick=\"storelink('".$row2[0]."', '".get_client_ip().
        "','".$_SESSION["username"]."')\"
        >".$row2[0]."</a></td><td>",$row2[1], "</td></tr>";}
    }
}
}

```

A4. Adding New Websites to Database

```

$result4=mysqli_query($con,"SELECT webex FROM
web_extension")or die("Wrong Query4");

while ($row4 = mysqli_fetch_row($result4))
{
    $extension=$row4[0];

    $url = "http://www.".$partofurl[0].$extension;
}

```

```

$ch = curl_init($url);

curl_setopt($ch, CURLOPT_HEADER, true);    // we want
headers

@curl_setopt($ch, CURLOPT_NOBODY, true);    // dont
need body

curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);
//TRUE to return the transfer as a string of the
return value of curl_exec() instead of outputting it
out directly.

curl_setopt($ch, CURLOPT_CONNECTTIMEOUT_MS, 3000);

$data = curl_exec($ch);

if($data !== false)

{
$title=@get_title($url);//$title='nothing';

//echo $url, " ", $data, "<br><br><br>";
$result5=mysqli_query($con,"SELECT COUNT(id) FROM
website;")or die("Wrong Query5");
$result6=mysqli_query($con,"SELECT MAX(id) FROM
website;")or die("Wrong Query6");

$row5 = mysqli_fetch_array($result5);

$row6 = mysqli_fetch_array($result6);
if(strpos($url,"http://")!==false) // Using !=
would not work as position is 0. The statement (0 != false)
evaluates to false.

{
$url=substr($url,7);}

elseif (strpos($url,"https://")!==false)

{
$url=substr($url,8);}

if($row5[0]==0)

{$result7=mysqli_query($con,"INSERT INTO website

```

```

values (1, '". $url.', '". $title.');" or die("Wrong
Query7.1");}

    else if ($row5[0]==$row6[0])

{$result7=mysqli_query($con,"INSERT INTO website values
('.$row5[0]+1).', '". $url.', '". $title.');" or die("Wrong
Query7.2");}

else

{$result8=mysqli_query($con,"SELECT id FROM website;") or
die("Wrong Query8");

    $idcheck=1;

    while ($row8=mysqli_fetch_row($result8))

    {

        if($row8[0]!=$idcheck)

        {

            $result7=mysqli_query($con,"INSERT INTO website values
('.$idcheck.', '". $url.', '". $title.');" or die("Wrong
Query7.3");
            break;}

        else

        {

            $idcheck=$idcheck+1;}}}

echo "<tr class='success'><td><a
onclick=\"storelink('". $url.', '".get_client_ip()."', '". $_S
SESSION["username"] . "')\"
>". $url."</a></td><td>",$title,"</td></tr>";}

    curl_close($ch);}

```


A5. Information Grabbing To Store with newly collected site

```
function get_title($url){
    $str = file_get_contents($url);
    if(strlen($str)>0)
    {
        preg_match("/\<title\>(.*?)\</title\>/i",$str,$title); //
        ignore case
        $title[1]=str_replace ("'", "\"", $title[1]);
        $title[1]=str_replace ('"', '\\"', $title[1]);
        if ($title[1]==' ' || $title[1]==null) {
            preg_match("/\<h1\>(.*?)\</h1\>/i",$str,$title);
            if ($title[1]==' ' || $title[1]==null) {
                preg_match("/\<h2\>(.*?)\</h2\>/i",$str,$title);
                if ($title[1]==' ' || $title[1]==null) {
                    preg_match("/\<h3\>(.*?)\</h3\>/i",$str,$title);
                    if ($title[1]==' ' || $title[1]==null) {
                        preg_match("/\<h4\>(.*?)\</h4\>/i",$str,$title);
                        if ($title[1]==' ' || $title[1]==null) {
                            preg_match("/\<p\>(.*?)\</p\>/i",$str,$title);
                        }
                    }
                }
            }
        }
    }
}
```

```

        }
    }
}

$title[1]=str_replace ('"',"\'",$title[1]);
$title[1]=str_replace ('\'','\"',$title[1]);
return $title[1];
}
}

```

A6. Data Acquisition for Processing

Phase 1:

```

echo "<tr class='success'><td><a
onclick=\"storelink('".$row0[0]."', '".$get_client_ip()."', '$_SESSION["username"]')."\"
>".$row0[0]."</a></td><td>",$row0[1], "</td></tr>";

```

Phase 2:

```

function storelink(s,t,u)
{
    window.open('processclick.php?site=' + s+
"&username=" + u+ "&userip=" + t, '_blank'); // <- This
is what makes it open in a new window.
}

```

Phase 3:

```

<?php
    if( isset($_GET["site"]) && isset($_GET["username"])
&& isset($_GET["userip"]))

```

```

{
    $z=$_GET["site"];
    $x=$_GET["username"];
    $y=$_GET["userip"];

    $con=mysqli_connect('localhost','root','') or die
('Connection Failed');

    $sel=mysqli_select_db($con,'thesisdemo') or die
('Database Connection Failed');

    $result1=mysqli_query($con,"SELECT id from
website where webadd='".$z."'");or die("Wrong
Query1");

    $row1 = mysqli_fetch_array($result1);

    $result3=mysqli_query($con,"SELECT * from
temptable where uname='".$x.'" and uip='".$y.'"
and id='".$row1[0].'");or die("Wrong Query3");

if(mysqli_num_rows($result3)==0)
{
    $result2=mysqli_query($con,"INSERT INTO temptable
values ('".$x."','".$y."','".$row1[0].'");or
die("Wrong Query2");
}

header("Location: http://".$z);}

?>

```

A7. Conversion of Temporary User Data to Raw Linked Data and Logging out a Session

```
<?php

    session_start();

    $nam=$_SESSION['username'];

    $ip=get_client_ip();

    $con=mysqli_connect('localhost','root','') or die
    ('Connection Failed');

    $sel=mysqli_select_db($con,'thesisdemo') or die
    ('Database Connection Failed');

    $result1=mysqli_query($con,"SELECT COUNT(DISTINCT tid)
    FROM transection;")or die("Wrong Query1");

    $result2=mysqli_query($con,"SELECT MAX(tid) FROM
    transection;")or die("Wrong Query2");

    $row1 = mysqli_fetch_array($result1);

    $row2 = mysqli_fetch_array($result2);

    echo $row1[0],"<br>",$row2[0],"<br>";

    $result3=mysqli_query($con,"SELECT id FROM temptable
    where uname='".$.$nam.'" and uip='".$.$ip.'";")or
    die("Wrong Query3");

    if (mysqli_num_rows($result3)>1) {

        while ($row3 = mysqli_fetch_row($result3))

            {

                if($row1[0]==0)

                    {
```

```

$result4=mysqli_query($con,"INSERT INTO
transaction values (1, ".$row3[0].");")or
die("Wrong Query4");

}

else if ($row1[0]==$row2[0]) {

$result5=mysqli_query($con,"INSERT INTO
transaction values
(".$row1[0]+1).", ".$row3[0].");")or
die("Wrong Query5");

}

else

{

$result6=mysqli_query($con,"SELECT tid FROM
transaction;")or die("Wrong Query6");

$idcheck=1;

while ($row6=mysqli_fetch_row($result6)) {

    if($row6[0]!=$idcheck)

    {

        $result7=mysqli_query($con,"INSERT
        INTO transaction values
        (".$idcheck.",' ".$row3[0]."');")or
        die("Wrong Query7");

        break;

    }

    else

    {$idcheck=$idcheck+1;}

}

}

}

```

```

    }

    $result8=mysqli_query($con,"DELETE FROM temptable
where uname='".$nam."' and uip='".$ip."'");or die("Wrong
Query8");

session_destroy();

header("Location: index.php");

?>

```

A8. Optimizing Linked Data by Mining (Using Apriori)

```

<?php

    ini_set('max_execution_time', 1000);

    $con=mysqli_connect('localhost','root','') or die
('Connection Failed');

    $sel=mysqli_select_db($con,'thesisdemo') or die
('Database Connection Failed');

    $result10=mysqli_query($con,"DELETE FROM fainalresult
WHERE 1")or die("Deleting old result Failed");

    $result=mysqli_query($con,"SELECT count(id) from
website")or die("Wrong Query");

    $resul=mysqli_query($con,"SELECT count(distinct tid)
from transection")or die("Wrong Query");

    $resu=mysqli_query($con,"SELECT count(tid) from
transection")or die("Wrong Query");

    $row = mysqli_fetch_row($result);

    $ro = mysqli_fetch_row($resul);

```

```

    $row = mysqli_fetch_row($resu);
    $total_website=$row[0];
    $valid1=0;
    $valid2=0;
    $invalid2=0;
    $k=0;
    $l=0;
    $cnt=0;
    $temp_com3=0;
    $tot_com3=0;

echo "<table class='table'><tr class='info'><th>Web
Address</th><th>Number of Occurrences</th></tr>";

for ($i=1; $i <$total_website ; $i++)
{
    $result1=mysqli_query($con,"SELECT count(webid) FROM
transection WHERE webid = $i")or die("Wrong Query");

    $row = mysqli_fetch_row($result1);
    $result2=mysqli_query($con,"SELECT webadd FROM website
WHERE id = $i")or die("Wrong Query");

    $row2 = mysqli_fetch_row($result2);

    $time_end = microtime(true);

    $time1 = $time_end - $time_start;

    if ($row[0]>1) {

echo "<tr class='success'><td>(<a
href='http://'",$row2[0],">",$row2[0],"</a>',"</td><td
align='center'>",$row[0],"</td></tr>";

    $item[$valid1]=$i;

    $valid1=$valid1+1;

```

```

    }

}

echo "<table class='table'><tr class='info'><th>Web
Address</th><th>Number of Occurrences</th></tr>";

//////////finding possible combination in 2nd
generation//////////

for ($i=0; $i < $valid1; $i++) {
    for ($j=1; $j < $valid1; $j++) {
        if($j>$i)
            {
                $item2[$k][0]=$item[$i];
                $item2[$k][1]=$item[$j];
                $k=$k+1;
            }
    }
}

//////////storing all transection in variable
length 2d array//////////

$tot_com2=($valid1*($valid1-1))/2;

$result39=mysqli_query($con,"SELECT max(tid) FROM
transection")or die("Wrong Query");//for checking total how
many transections

$row39 = mysqli_fetch_row($result39);

for ($j=1; $j <=$row39[0] ; $j++)
{
$result49=mysqli_query($con,"SELECT count(webid) FROM
transection WHERE tid = '$j'")or die("Wrong Query");// for
checking how many elements in each transection

```



```

$result59=mysqli_query($con,"SELECT webid FROM transection
WHERE tid = '$j'")or die("Wrong Query");

$row49 = mysqli_fetch_row($result49);

$a=array();

$a[$j-1] = array();

for($r=0; $r< $row49[0]; $r++){

    $row59 = mysqli_fetch_row($result59);

    $a[$j-1][$r] = $row59[0];

    }

}

//////////finding occurrence of combination in 2d
array//////////

for ($z=0; $z < $tot_com2; $z++) {

$total_item[$z]=0;

$result3=mysqli_query($con,"SELECT distinct tid FROM
transection")or die("Wrong Query");

while ($row3 = mysqli_fetch_row($result3)) {

//echo $row3[0]," ",$item2[$z][0]," ",$item2[$z][1];//every
value used in query ok

$result4=mysqli_query($con,"SELECT count(webid) FROM
transection WHERE (tid=".$row3[0]." and
(webid=".$item2[$z][0]." or webid=".$item2[$z][1]."))")or
die("Wrong Query");//but still this is not working

$row4 = mysqli_fetch_row($result4);

if ($row4[0]>1) {

    $total_item[$z]=$total_item[$z]+1;

    }

}

```

```

    }

$result20=mysqli_query($con,"SELECT webadd FROM website
WHERE id = ".$item2[$z][0])or die("Wrong Query");

$row20 = mysqli_fetch_row($result20);

$result21=mysqli_query($con,"SELECT webadd FROM website
WHERE id = ".$item2[$z][1])or die("Wrong Query");

$row21 = mysqli_fetch_row($result21);

    if ($total_item[$z]>1) {

        $time_end = microtime(true);

        $time2 = $time_end - $time_start;

        echo "<tr class='success'><td><a
href='http://",$row20[0],"'>",$row20[0],"</a>",<a
href='http://",$row21[0],"'>",$row21[0],"</a>",</td><td
align='center'>",$total_item[$z],"</td></tr>";

        $validitem2[$valid2][0]=$item2[$z][0];

        $validitem2[$valid2][1]=$item2[$z][1];

//echo $validitem2[$valid2][0],"
",$validitem2[$valid2][1],"<br>";

        $valid2=$valid2+1;

    }

else

{

//echo "<tr
class='danger'><td>(", $row20[0],"), (", $row21[0],")</td><td
align='center'>",$total_item[$z],"</td></tr>";

        $invaliditem2[$invalid2][0]=$item2[$z][0];

```

```

        $invaliditem2[$invalid2][1]=$item2[$z][1];
//echo $invaliditem2[$invalid2][0], "
", $invaliditem2[$invalid2][1], "<br>";

        $invalid2=$invalid2+1;

        }

    }

//////////finding total combination possible in 3d
array with my own formula//////////
for ($i=$valid1; $i > 1; $i--) {

        $temp_com3=($i-1)*($i-2)/2;

        $tot_com3=$tot_com3 + $temp_com3;

    }

//////////finding occurrence of combination in 3d
array//////////
for ($z=0; $z < $l; $z++) {

        $total_item3d[$z]=0;

        $result3=mysqli_query($con, "SELECT distinct tid FROM
        transection") or die("Wrong Query");

        while ($row3 = mysqli_fetch_row($result3)) {

            $result4=mysqli_query($con, "SELECT count(webid) FROM
            transection WHERE (tid=".$row3[0]. " and
            (webid=".$item3[$z][0]. " or webid=".$item3[$z][1]. " or
            webid=".$item3[$z][2]. ")))" or die("Wrong Query");//but
            still this is not working

            $row4 = mysqli_fetch_row($result4);

            if ($row4[0]>2) {

                $total_item3d[$z]=$total_item3d[$z]+1;

            }

        }

    }

```

```

    }

$result20=mysqli_query($con,"SELECT webadd FROM website
WHERE id = ".$item3[$z][0])or die("Wrong Query");

$row20 = mysqli_fetch_row($result20);

    $result21=mysqli_query($con,"SELECT webadd FROM
website WHERE id = ".$item3[$z][1])or die("Wrong Query");

$row21 = mysqli_fetch_row($result21);

    $result22=mysqli_query($con,"SELECT webadd FROM
website WHERE id = ".$item3[$z][2])or die("Wrong Query");

$row22 = mysqli_fetch_row($result22);

echo "<tr><td>(<a
href='http://",$row20[0],">",$row20[0],"</a>",<a
href='http://",$row21[0],">",$row21[0],"</a>",<a
href='http://",$row22[0],">",$row22[0],"</a>",</td><td
align='center'>",$total_item3d[$z],"</td></tr>";

    }

    $time_end = microtime(true);

    $time3 = $time_end - $time_start;

//////////finding occurrence of combination in 3d
array//////////

for ($z=0; $z < $l; $z++) {

    $total_item3d[$z]=0;

    $result3=mysqli_query($con,"SELECT distinct tid
FROM transection")or die("Wrong Query");

    while ($row3 = mysqli_fetch_row($result3)) {

        $result4=mysqli_query($con,"SELECT
count(webid) FROM transection WHERE (tid=".$row3[0].
and (webid=".$item3[$z][0]. or webid=".$item3[$z][1]. or
webid=".$item3[$z][2].)")or die("Wrong Query");//but
still this is not working

```

```

$row4 = mysqli_fetch_row($result4);

    if ($row4[0]>2) {
        $total_item3d[$z]=$total_item3d[$z]+1;
    }
}

    $result20=mysqli_query($con,"SELECT webadd
FROM website WHERE id = ".$item3[$z][0])or die("Wrong
Query");

    $row20 = mysqli_fetch_row($result20);

    $result21=mysqli_query($con,"SELECT webadd FROM
website WHERE id = ".$item3[$z][1])or die("Wrong Query");

    $row21 = mysqli_fetch_row($result21);

    $result22=mysqli_query($con,"SELECT webadd FROM
website WHERE id = ".$item3[$z][2])or die("Wrong Query");

    $row22 = mysqli_fetch_row($result22);

    if ($total_item3d[$z]>1) {

        echo "<tr class='success'><td><a
href='http://",$row20[0],"'>",$row20[0],"</a>,"") (<a
href='http://",$row21[0],"'>",$row21[0],"</a>,"") (<a
href='http://",$row22[0],"'>",$row22[0],"</a>,"")</td><td
align='center'>",$total_item3d[$z],"</td></tr>";

        $result9=mysqli_query($con,"INSERT INTO fainalresult
values
('".$item3[$z][0].",".$item3[$z][1].",".$item3[$z][2].");")o
r die("Insering new data failed");

    }

}

```

```

        echo "Total number of different website's:-
><strong>",$total_website,"</strong><br>";

        echo "Total number of set:-
><strong>",$ro[0],"</strong><br>";

        echo "Total websites in transection:-
><strong>",$rw[0],"</strong><br>";

        echo "1st degree process time needed
=" .round(($time1), 2)." seconds<br>";

        echo "2nd degree process time needed
=" .round(($time2), 2)." seconds<br>";

        echo "3rd degree process time needed
=" .round(($time3), 2)." seconds<br>";

        mysqli_close($con);

?>

```

A9. Providing Benefit to User (1 example)

```

$fnresult0=mysqli_query($con,"SELECT linkedata1 from
fainalresult WHERE linkedata1=".$fnrow0[0]." or
linkedata2=".$fnrow0[0]." or linkedata3=".$fnrow0[0]."
UNION SELECT linkedata2 from fainalresult WHERE
linkedata1=".$fnrow0[0]." or linkedata2=".$fnrow0[0]." or
linkedata3=".$fnrow0[0]." UNION SELECT linkedata3 from
fainalresult WHERE linkedata1=".$fnrow0[0]." or
linkedata2=".$fnrow0[0]." or
linkedata3=".$fnrow0[0].";") or die(mysqli_error($con));

if (mysqli_num_rows($fnresult0)>0)
{
echo "<tr><td colspan='3' style='padding:20px 80px;
border:2px solid #DFF0D8;'><ul>";

while ($fnrow0 = mysqli_fetch_row($fnresult0))

```

```
{
if ($fnrow0[0]!=$fnrow00[0])
{
$fnresult01=mysqli_query($con,"SELECT webadd,univ_name from
website where id='".$fnrow0[0]."'");or die("Final result
Wrong Query01");
$fnrow1 = mysqli_fetch_row($fnresult01);
echo"<li><a
href='#'>".$fnrow1[0]."</a> ( ".$fnrow1[1].")</li>";
}
}
echo "</ul></tr>";
}
```

Full Code available at: <https://goo.gl/UrD5Xz>