Zooming Prediction and Hyperlink Structure on Web

Madhina.H	**Mrs. S. Revathy, M.E.,**
ME – Software Engineering	Computer Science and Engineering
GKM College of Engineering and Technology	GKM College of Engineering and Technology
Chennai, India	Chennai India

Abstract-In recent years, there is a rapid development of the communication and computer technology. The web search is most important for the life style. It carryout the user interests. A novel technique to learn user profiles from user search histories. The user profiles are then used to improve retrieval effectiveness in web search. By utilizing the selected categories as a context for the query, a search engine is likely to return documents that are more suitable to the user. Unfortunately, a category hierarchy shown to a user is usually very large, and as a result, an ordinary user may have difficulty in finding the proper paths leading to the suitable categories. Furthermore, users are often too impatient to identify the proper categories before submitting his/her queries. An alternative to browsing is to obtain a set of categories for a user query directly by a search engine. However, categories returned from a typical search engine are still independent of a particular user and many of the returned categories do not reflect the intention of the searcher. To solve this problem a zooming system and tag system is proposed. The project also discuss the website's new search engine. And the updation of the product or user interest in the web search.

Index Terms—Web search, full zoom, tags, feature selection

I.   INTRODUCTION

The Web has become the most important way to spread information. Archiving the Web is thus crucial to preserve some useful information for future generations of historians, researchers, writers, or citizens. Most of the time, Web archiving is performed by Web crawlers (bots) that capture Web pages. To update archives, crawlers have to regularly revisit Web pages, but they generally do not know if or when changes appeared. A web search engine is a software system that is designed to search for information on the World Wide Web. The search results are generally presented in a line of results often referred to as search engine results pages (SERPs). The information may be a specialist in web pages, images, information and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintained only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

II.   OBJECTIVE

A user may associate one or more categories to his/her query manually. For example, a user may first browse a hierarchy of categories and select one or more categories in the hierarchy before submitting his/her query. By utilizing the selected categories as a context for the query, a search engine is likely to return documents that are more suitable to the user. Unfortunately, a category hierarchy shown to a user is usually very large, and as a result, an ordinary user may have difficulty in finding the proper paths leading to the suitable categories. Furthermore, users are often too impatient to identify the proper categories before submitting his/her queries. An alternative to browsing is to obtain a set of categories for a user query directly by a search engine. However, categories returned from a typical search engine are still independent of a particular user and many of the returned categories do not reflect the intention of the searcher. To solve these problems, we propose a two-step strategy to improve retrieval effectiveness.

Visit your favorite news page and read the caption under the picture or view the picture itself in a size you can see. An elegant new zoom feature lets you swoop in and see entire web pages.It allow cropping and enlarging of a captured image, in order to emulate the effect of a longer focal length zoom lens (narrower angle of view). This is commonly known as digital zoom and produces an image of lower optical resolution than optical zoom. Exactly the same effect can be obtained by using digital image processing software on a computer to crop the digital image and enlarge the cropped area.

There are many possible designs for zoom lenses, the most complex ones having upwards of thirty individual lens elements and multiple moving parts. Most, however, follow the same basic design. Generally they consist of a number of individual lenses that may be either fixed, or slide axially along the body of the lens. While the magnification of a zoom lens changes, it is necessary to compensate for any movement of the focal plane to keep the focused image sharp. This compensation may be done by mechanical means (moving the complete lens assembly while the magnification of the lens changes), or optically (arranging the position of the focal plane to vary as little as possible while the lens is zoomed).

Logs are often created by software developers to aid in the debugging of the operation of an application. The syntax and semantics of data within log messages are usually application or vendor-specific. Terminology may also vary; for example, the authentication of a user to an application may be described as a login, a logon, a user connection or authentication event. Hence, log analysis must interpret messages within the context of an application, vendor, system or configuration in order to make useful comparisons to messages from different log sources.

These techniques generally fall into the following categories: data modification techniques, cryptographic methods statistical techniques for disclosure and inference control, query auditing methods, randomization and perturbation-based techniques.

This edited volume also contains surveys by distinguished researchers in the privacy field. Each survey includes the key research content as well as future research directions of a particular topic in privacy.

We propose a new solution by integrating the advantages of both these techniques with the view of minimizing information loss and privacy loss. By making use of cryptographic techniques to store sensitive data and providing access to the stored data based on an individual's role, we ensure that the data is safe from privacy breaches.

III.   SYSTEM ANALYSIS

Current web search engines are built to serve all users, independent of the special needs of any individual user. Personalization of web search is to carry out retrieval for each user incorporating his/her interests. The system we going to develop is very user friendly. The accuracy of data is provided by proper validation of inputs. A selection-based search system allows the user to search the internet for more information about any keyword or phrase contained within a document or webpage in any software application on his desktop computer using the mouse. The retrieval of necessary information is very easy and fast. Manual options are reduced to maximum possible extent. Backups of data are maintained. The processing of data is faster when compared to the existing system. On-line access is also possible and good to use. Furthermore, users are often too impatient to identify the proper categories before submitting his/her queries. An alternative to browsing is to obtain a set of categories for a user query directly by a search engine. However, categories returned from a typical search engine are still independent of a particular user and many of the returned categories do not reflect the intention of the searcher. To solve these problems, a two-step strategy to improve retrieval effectiveness called tagging and zooming.

☐ Improve your competitive edge
☐ Expand customer base and target audience
☐ Boost sales / Increase your return on investment
☐ save time and money

IV.   SYSTEM DESCRIPTION

The system is designed for the web search and the tagging system. Logs are often created by software developers to aid in the debugging of the operation of an application. The syntax and semantics of data within log messages are usually application or vendor-specific.
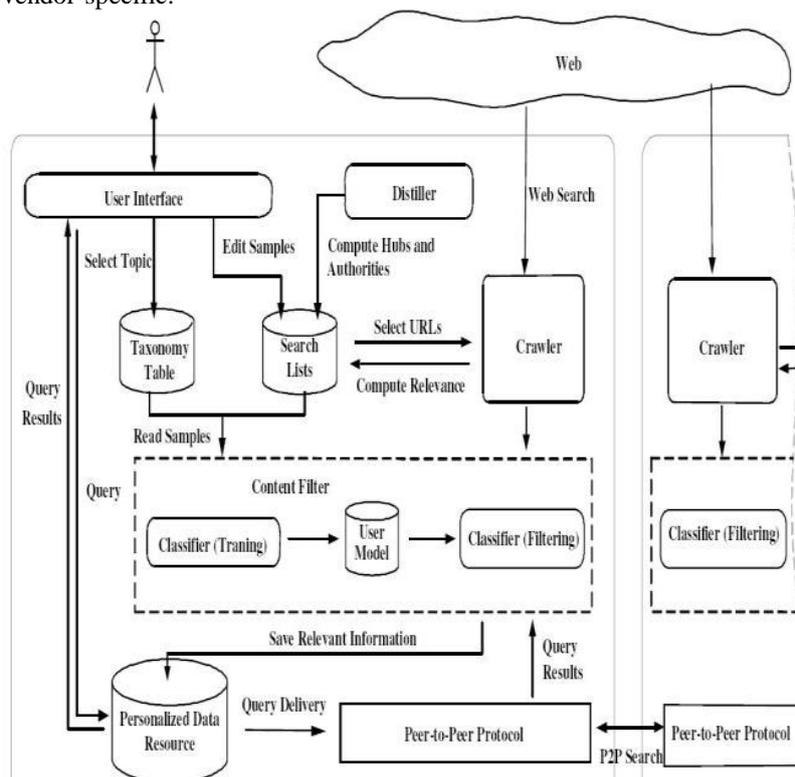

Fig 1 System Architecture

Terminology may also vary; for example, the authentication of a user to an application may be described as a login, a logon, a user connection or authentication event. Hence, log analysis must interpret messages within the context of an application, vendor, system or configuration in order to make useful comparisons to messages from different log sources.

V. FEASIBILITY STUDY

Economic analysis is the most frequently used method for evaluating the effectiveness of a new system. More commonly known as cost/benefit analysis, the procedure is to determine the benefits and savings that are expected from a candidate system and compare them with costs. If benefits outweigh costs, then the decision is made to design and implement the system. An entrepreneur must accurately weigh the cost versus benefits before taking an action. Cost Based Study: It is important to identify cost and benefit factors, which can be categorized as follows: 1. Development costs; and 2. Operating costs. This is an analysis of the costs to be incurred in the system and the benefits derivable out of the system. Time Based Study: This is an analysis of the time required to achieve a return on investments. the benefits derived from the system. The future value of a project is also a factor. Is a measure of how well a proposed system solves the problems, and takes advantages of the opportunities identified during scope definition and how it satisfies the requirements identified in the requirements analysis phase of system development. The assessment is based on an outline design of system requirements in terms of Input, Processes, Output, Fields, Programs, and Procedures. This can be quantified in terms of volumes of data, trends, frequency of updating, etc. in order to estimate whether the new system will perform adequately or not. Technological feasibility is carried out to determine whether the company has the capability, in terms of software, hardware, personnel and expertise, to handle the completion of the project.

MODULE DESCRIPTION

There are two types of users Sensitive and Non Sensitive objects are involved in this project to accessing the files in the server. Sensitive objects are higher end users such as project manager that predetermined by Admin and accessed only according to role permission.

MODULE 1: AWESOME BAR

A quick way to get to the sites you love—even the ones with addresses you only vaguely remember. Type in term into location bar (the Awesome Bar) and the auto complete function includes possible matching sites from your browsing history, as well as sites you've book marked and tagged in a drop down.

MODULE 2: TAGS

When you enter "news" into the location bar, both sites will be shown as results. A single site can have multiple tags, and there's no limit to the number of tags you can create. You may not remember the exact name of a site, but with a tag, you'll be able to find it in a way that makes sense to you.

MODULE 3: FULL ZOOM

Visit your favorite news page and read the caption under the picture—or view the picture itself in a size you can see. An elegant new zoom feature lets you swoop in and see entire web pages. They scale in the way you'd expect them to, with all the elements of a page's layout expanding equally, so you can zero in on what matters.

MODULE 4: SMART KEYWORDS

Search the Web in record time with smart keywords. With a few clicks you can assign keywords to search engines, and then simply enter your key and search words in the location bar. With this feature, typing "book home building" can search Amazon.com, and take you right to books about home building without ever pausing at the homepage.

MODULE 5: LATEST SEARCH HITS

In the 'Latest search terms' user area the latest search engine hits are displayed in real-time, along with landing page and search engine information. Be careful, as you can get addicted to checking this section. Includes ranking information.

MODULE 6: USER PROFILING

In the Web domain, user profiling is the process of gathering information specific to each visitor, either explicitly or implicitly. A user profile includes demographic information about the user, her interests and even her behavior when browsing a Web site. This information is exploited in order to customize the content and structure of a Web site to the visitor's specific and individual needs.

MODULE 7: LOG ANALYSIS

Web log analysis software (also called a web log analyzer) is a simple kind of Web analytics software that parses a log file from a web server, and based on the values contained in the log file, derives indicators about who, when, and how a webserver is visited. Usually reports are generated from the log files immediately, but the log files can alternatively be parsed to a database and reports generated on demand.

MODULE 8: CONTENT MANAGEMENT

This is the process of classifying the content of a Web site in semantic categories in order to make information retrieval and presentation easier for the users. Content management is very important for Web sites whose content is increasing on a daily basis, such as news sites or portals.

MODULE 9: CAMPAIGN PLANNING, EXECUTION AND MANAGEMENT:

Easily create, view and edit campaigns, and share and reuse campaigns. Ensure proper sign-offs before campaigns are executed with approval and automated notification processes. Allow remote or distributed users to open, review, approve and edit campaigns and communications, and to publish campaign reports, via a Web user interface. View campaign-related creative collateral – brochures, inserts, e- mails, banner ads and other documents.

SYSTEM TESTING:
TESTING TOOL
BUGZILLA is a web base system using for bug- or issue- tracking in software developing under quality assurance. In sense of a Bug tracking system, it is a tool, which allows individual or groups of developers effectively to keep track of unsettled problems with their product. Terry Weissman originally wrote Bugzilla in 1998 using a programming language called TCL, to replace a rudimentary bug-tracking database used internally by Netscape Communications. Terry

later ported Bugzilla to Perl from TCL, and in Perl it remains to this day. Bugzilla quickly became a favorite of the open- source crowd (with its genesis in the open-source browser project, Mozilla). It is now the de-facto standard defect- tracking system against which all others are measured. Potential exists in the code to turn Bugzilla into a technical support ticket system, task management tool, or project management tool, Bugzilla's developers have chosen to focus on the task of designing a system to track software defects.

Mandated design requirements include: ☐ The ability to run on freely available, open source tools. While Bugzilla development includes work to support commercial databases,

tools, and operating systems. ☐ The maintenance of speed and

efficiency at all costs. One of Bugzilla's major attractions to developers is its lightweight implementation and speed, so calls into the database are minimized whenever possible, data fetching is kept as light as possible, and generation of heavy is

avoided. ☐ Tickets. For instance, Mozilla.org and the project

use it to track feature requests as well. In this case items (called bugs) can be submitted by anybody, and will be assigned to a particular developer Various status updates for each bug are allowed, together with user notes and bug examples.

Bugzilla boasts many advanced features. These include:

-configurable email notifications of bug changes ull change history
-bug dependency tracking and graphing

-based, granular security schema
-audited, and runs under Perl's taint mode
-end

ML, email and console interfaces interface

between versions ☐ Extensive configurability ☐ Smooth upgrade pathway

☐ Powerful searching
☐ User
☐ F
☐ Inter
☐ Excellent attachment management
☐ Integrated, product
☐ Fully security
☐ A robust, stable RDBMS back
☐ Web, X
☐ Completely customisable and/or localisable web user

UNIT TESTING
It composes the set of tests performed by the individual
programmer prior to integration of the larger system. It first focuses on the modules independent of one another to locate errors. The format is checked before the records get added in the database similarly records were deleted and checked they are removed from the database.

Because some classes may have references to other classes, testing a class can frequently spill over into testing another class. A common example of this is classes that depend on a database: in order to test the class, the tester often writes code that interacts with the database. This is a mistake, because a unit test should usually not go outside of its own class boundary, and especially should not cross such process/network boundaries because this can introduce unacceptable performance problems to the unit test-suite. Crossing such unit boundaries turns unit tests into integration tests, and when test cases fail, makes it less clear which component is causing the failure.

Instead, the software developer should create an abstract interface around the database queries, and then implement that interface with their own mock object. By abstracting this necessary attachment from the code (temporarily reducing the net effective coupling), the independent unit can be more thoroughly tested than may have been previously achieved. This results in a higher quality unit that is also more maintainable.

ACCEPTANCE TESTING
☐ Very often done by the Users.
☐ it is a type of testing to test whether it works within the

defined constraints.

☐ testing to ensure that the system meets the need of the organization and the end user/ Admin.

INPUT DESIGN

Input design of a system is preparation of input. This input design stage is necessary for successful development and implementation of the system. Input design is the process of converting user-originated input to a computer based format. Input data are collected and organized into similar group of data the goal of, input data is to make data entry as easy, logical and free from errors as possible. When the design input has been reviewed and the design input requirements are determined to be acceptable, an iterative process of translating those requirements into a device design begins. The first step is conversion of the requirements into system or high-level specifications. Thus, these specifications are a design output. Upon verification that the high-level specifications conform to the design input requirements, they become the design input for the next step in the design process, and so on. This basic technique is used repeatedly throughout the design process. Each design input is converted into a new design output; each output is verified as conforming to its input; and it then becomes the design input for another step in the design process. In this manner, the design input requirements are translated into a device design conforming to those requirements. The web site users are selecting the product and the tags are implemented. The full zoom concept is implemented after the relogin of the user.

VI.   OUTPUT DESIGN

One of the most important features of an information system for users in the output it procedures. Output is the information delivered to user through the information system. This design is essential to gain the user acceptance. Efficient Output design should improve the system relationship with the user. The concept of zooming prediction is to give the updation of the entries of the users to the admin and the updation of the product ratings to the user. The tags are implemented for the remembrance of the user. These are fulfilled their requirements when the user again login the website. He can update everything automatically by using the web site
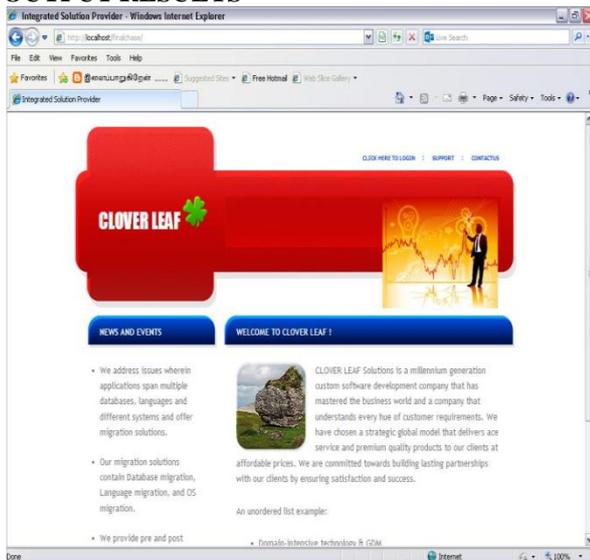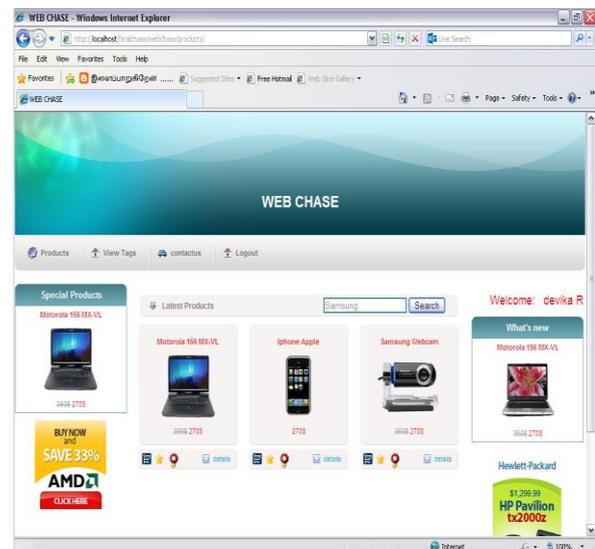
OUTPUT RESULTS

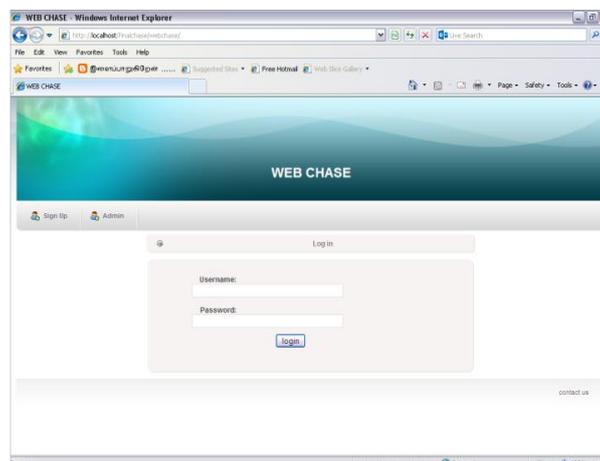**Fig 2 Welcome Page**

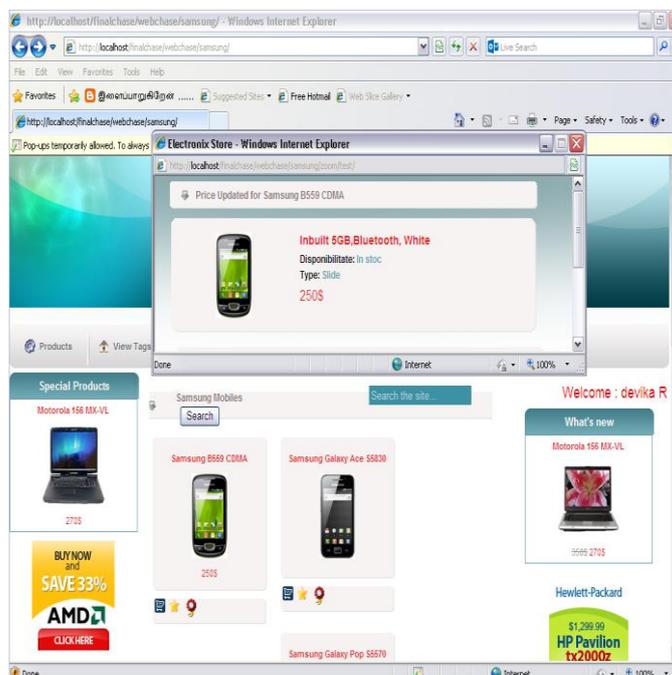**Fig 3 Login Page**

Fig 4 Tagging Page

Fig 5 Zooming Page

VII.   FUTURE ENHANCEMENTS

Further enhancement can also be made in the system user requirement. According to user requirements, add some more modules and facilities in future. This project has been developed to satisfy the web users. One of the most important features of an information system for users in the output it procedures. Output is the information delivered to user through the information system. This design is essential to gain the user acceptance. Efficient Output design should improve the system relationship with the user. Cost of the time when we print records in FileMaker, the output of every record appears pretty much the same, with similar formatting and design. But what if we want records to print with a completely different output design depending on the type of record? In this article we detail some of the various ways in which this can be achieved, and the pros and cons of each method. Each has been formatted as we wish for the question type to appear. Now it is simply a case of trying to hide the two designs that do not apply to the type of question for any given record. When all of the databases are stacked on top of each other, a cool thing happens. The two portals that do not show a record will not only show nothing, but they completely disappear from the output, it's as if they aren't on the layout. This leaves the one remaining portal that is showing contents to be the only portal visible on that record. A container field has been added to the table. The idea is that we populate this container field with an image that looks like the question design we want to output. To do this, we basically go to one of the three layouts, take a picture of the layout, then go back to the container and paste in the results. From software point of view the system can be enhanced using any other backend tool, so that the system can also be further enhanced.

VIII.   CONCLUSION

The project proposes the new ideas of web site .The web site proposes the tagging and full zooming system to user. It introduces the new way to achieve the web search and retrieval of the updation. And it produces the new kind of search engine. It must fulfill the user query by automatically updating the users wish that is tagged before. The company has been quite successful in improving the efficiency in software handling. The company has been fairly successful in achieving considerable savings in manpower and man-hours involved in software handling. This has helped the company to reduce the manual cost significantly. So, as to take the advantage of the developments in the technology and to maintain a higher level of operation.

REFERENCES

[1]   Bricolage: A Structured-Prediction Algorithm for Example-Based Web Design, Jerry O. Talton, Salman Ahmad.
[2]   A Quantitative Method for Quality Evaluation of Web Sites and Applications, Luis Olsina, Gustavo Rossi
[3]   The Structure of Broad Topics on the Web", Soumen Chakrabarti, David M. Pennock
[4]   Waiting times in quality of experience for web based services , S. Egger, T. Hossfeld
[5]   WEST: A Web Browser for Small Terminals, Jussi Karlgren, Ivan Bretan
[6]   A Model-Based Approach for Supporting Offline Interaction with Web Sites Resilient to Interruptions, Felix Albertos Marco, José Gallud, Victor Penichet.
[7]   Real-Time Network Traffic Prediction based on a Multiscale Decomposition, Guoqiang Mao
[8]   Exploring Web Scale Language Models for Search Query Processing, Jian Huang, Jianfeng Gao