

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

*IJCSMC, Vol. 4, Issue. 3, March 2015, pg.753 – 757*

### **RESEARCH ARTICLE**

# Complete Human Portal (emarathi.com)

Chanakya N Bagul<sup>1</sup>

Final Year, Computer Engineering, S.V.I.T. Nashik, Pune University, India

<sup>1</sup> [sfchanakya@gmail.com](mailto:sfchanakya@gmail.com)

---

*Abstract— eMarathiPortal.com is a portal as stated in its domain name. This portal is focused only on Marathi language and same community who deals in Marathi. There are different views while developing eMarathiPortal.com. Initially by developing this portal we want to support and promote all kinds of writers, poets and all the people who blogs in their respected fields. Also other motto while developing this portal is to generate some revenue by running portal. Advertisement is the best way to fulfill economic demand.*

*Keywords—emarathi portal, Data mining, Jokes, Poems, Joomla, Administrator*

---

## INTRODUCTION

This website will contain all the stated data in regional language. The objective for this web portal is to promote various writers, poets and different professionals and also to generate revenue through different advertisements. The matter for this web portal will be collected from various professional sources.

All the information gathers in portal by using web mining. Web mining consists of three aspects: Web content mining, Web build mining, and web control mining. The most useful application of web mining is targeted billing. Constant mining is the process of applying data mining techniques to a constant database for the purposes of discovering the correlation relationships that exist among an ordered list of events. An important application of constant mining techniques is web usage mining, for mining web log door, where the arrays of web page accesses made by different web users over a cycle of time, through a server, are filed. This paper proposes web log constant pattern mining using A priori all algorithms. We called as A priori all Web log Mining. The experiment will be conducted base on the idea of A priori all algorithm, which first stores the original web access sequence database for storing non-constant data. The experimental result will be given with analysis on further refinement.

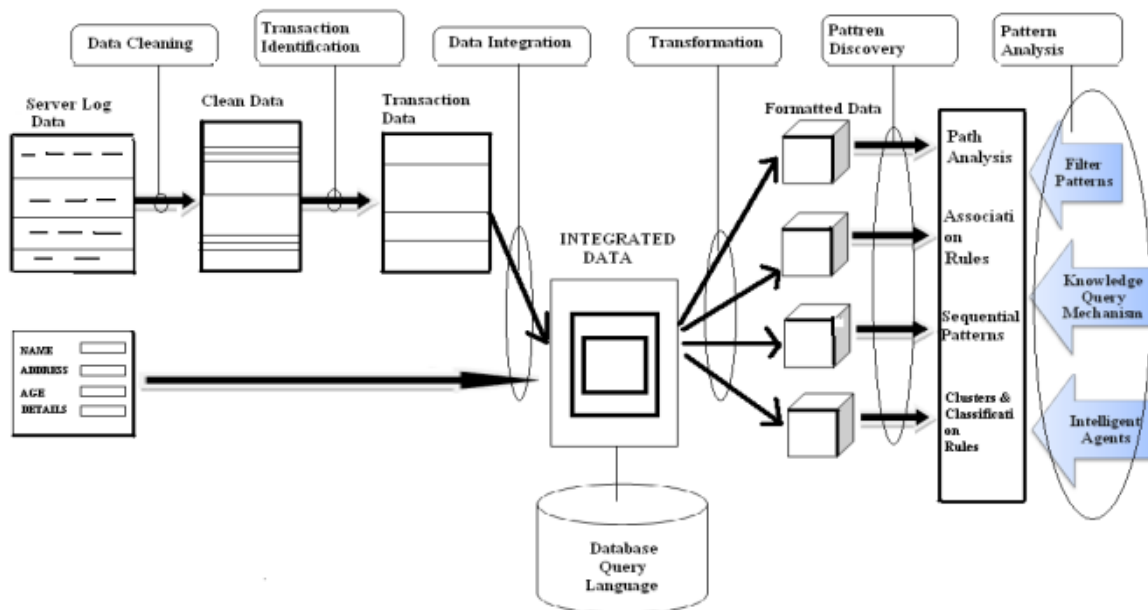


Fig .1 Data mining architecture

**EXISTING SYSTEM**

Existing web portal in nothing but just a bunch of huge data collected from resources and placed in static manner. Now client need to make a CMS (Content Management System) for same portal so writers and other resources can post articles on web portal by their own by using CMS. Also client want to promote this portal on google by using Search Engine Optimization (SEO) .

**SCOPE OF PROPOSED SYSTEM**

Whole site is nothing but Website that mainly for Users, Administrator as well as Client. The process started which consist of many steps like Registration of User, their user details, their ideas. At administrator side, he will decide overall changes in the site. Also He will check all the documents authenticate the entries of every registered user

Every process is interdependent. So output of one process fed as input to other. The schedule is decided first System requests to site. The project aims at handling Registration & time saving. That everyone gets updates & everyone share its own ideas easily & within time. Site gives their details and rules through this system.

**PROBLEM STATEMENT AND IMPLEMENTATION**

eMarathiPortal.com is developed in CMS framework. We have used Joomla! For same.

**Joomla!** is an award winning Content Management System. It is developed in PHP & MySQL. Joomla! itself have lot more functionalities in its core development, like User Registration, User Blocks (Modules), Contact form, Menu system, frontend & backend content editing and lot more to mention... Joomla! uses external developed functionalities to enhance its working along with cool and sleek designing. Such functionalities are separated in four different parts; **Template**; it's a design with updated W3C i.e. its responsive, user friendly, accessible. **Plug-in**; these are little core enhancements like Image gallery in content. **Components**; these are web applications to support this system in more accessible manner like Social component to convert this portal into Social networking site. **Modules**; these are user blocks supported by different functions like image show, login, footer, header and many more... We have used this system till it's extinct to make it usable in our ways.

By using different third party extensions we have implemented following systems in portal.

## Memberships

We have allotted many types of user access levels to filter data by checking it manually.

1. **Registered User** is the user who can access data which is accessible only to registered user and above this level. Such data will not be available publicly.
2. **Writer** is the user access level which can submit their data into respective category which will later make available for public after moderation.
3. **Publisher** is the user who can read data submitted by **Writer** access level, publisher will read content and if found proper, he will enable it. **Publisher** will not be able to create or modify data. If found improper content, publisher will forward such data to editor.
4. **Editor** is the user access level where filtered data is forwarded from **publisher**. Received data may have some improper words, spelling mistakes or grammatical mistakes to be corrected by **Editor**.
5. **Manager** is the user who has some more rights than editor and publisher. Manager has access to all above stated user's content.

Manager do have right to enable or disable Registered users.

6. **Moderator** is the user who has all rights regarding above all users. Moderator has full control over content and accounts of Registered User, Publisher, Editor and Manager. Manager will also work on some core features like modules, photo galleries, video galleries, contacts, polls and some other functionality.
7. **Administrator** is the user who has full rights same as moderator and partially access to backend control panel for further website development. Administrator can access all extensions i.e. Components, Modules, Plug-ins; but administrator cannot install any of the stated extension type.
8. **Super Administrator** name itself explains all. This user has full access same as all the above users. Also this user is the only user who is also a developer of same portal. This user can place banner ads and ads in portal blocks to generate revenue.

**Implementation Using Data Mining**

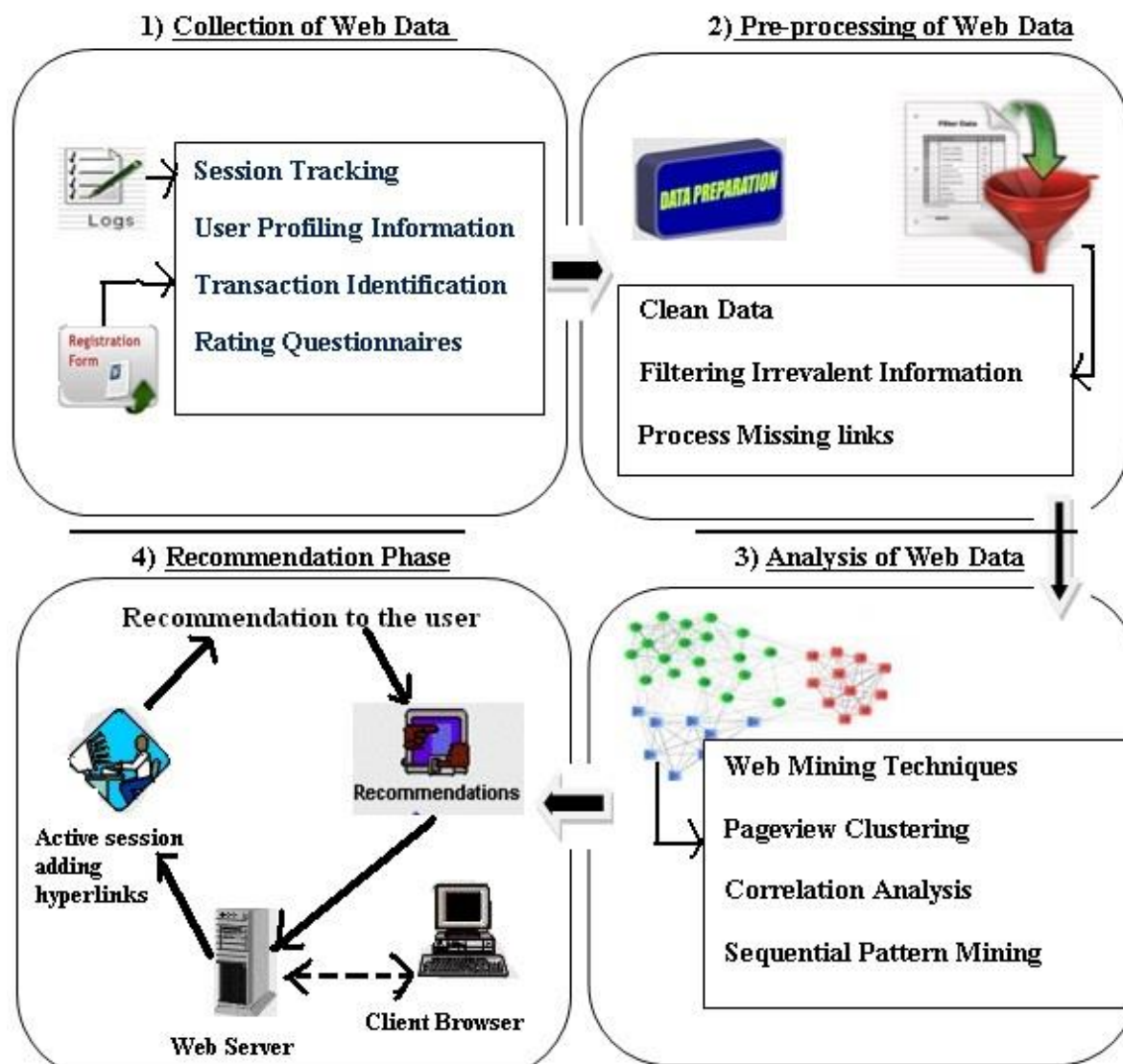


Fig .2 Web mining personalization process

**RELATED WORK:-**

Web-usage mining (WUM), an emergent domain in web mining that has greatly concerned both academia and industry in recent years. One of many possible applications of Web Usage mining, which is the action of applying data mining techniques to the discovery of usage patterns from Web data, targeted towards various applications. WUM is the process of discovering and interpreting patterns of user access to web information systems by mining the data collected from user interactions with the system. A typical WUM system consists of two tiers: tracking, in which user interactions are captured and acquired, and analysis, in which user access patterns are discovered and interpreted by applying typical data-mining techniques to the acquired data. There are three main tasks for performing WUM— pre-processing, figure discovery and figure analysis. As below:-

**Pre-processing:** It is generally used as groundwork of data mining method, data pre-processing cleaned/filtered the raw data to eliminate outliers or irrelevant items, alignment alone page accesses into semantic units for the purpose of the user. The different types of pre-processing in WUM are— usage, content, and structure pre-processing.

**Pattern Discovery:** In this, WUM can be able to unearth patterns in server logs and carried out only on samples of data. Interpretation and estimation of results be done on bits of data. The various pattern analysis methods are— Analytical Analysis, Association Rules, Clustering, share, Sequential Patterns, and Belief Modelling.

**Pattern Analysis:** The need behind pattern analysis is to filter out uninteresting rules or patterns from the set found in the pattern discovery phase. Most common form of pattern analysis consists of a knowledge query apparatus such as SQL. Content and structure information can be used to filter out patterns containing pages of a certain usage type, content type, or pages that match a certain hyperlink structure.

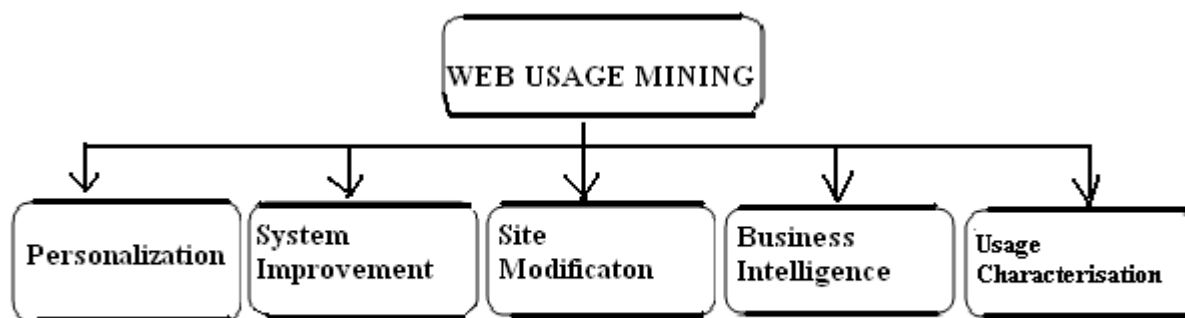


Fig.3 Web uses mining

**CONCLUSION**

We have taken view that Web personalization is appliance of data mining and therefore must be supported during the various phases of a typical data mining cycle. We defined framework for web personalization expert based of web mining and co-function filtering techniques. We adopted joint filtering technique combined with association rule mining technique, especially a priori algorithm respectively to associate the usage pattern of clients in particular website. The main limitation of Apriori algorithm is that the candidate set generation is high, particularly when immense number of the figure or long patterns continue. The main limitations of FP-growth algorithm is the explosive quantity of lacks a good candidate generating method. Future research can combination FP-Tree with Apriori candidate generation method to solve disadvantages of both apriori and FP-growth. The new algorithm will reduce the storage space, correct efficiency and accuracy of the algorithm within.

**References:**

1. Different professionals as per content.
2. [www.facebook.com](http://www.facebook.com)
3. [www.wikipedia.com](http://www.wikipedia.com)
4. Huge data mining will take place in this project... various methodologies will be followed to gather data.
5. Jiawei Han, Micheline Kamber, —Data mining big ideas and techniques!, Elsevier Inc., Second copy, San Francisco, 2006.
6. Charalampos Vassiliou, Dimitrios Stamoulis, Anastasias, —Creating Adaptive Web Sites Using Personalization Techniques: A Unified, Integrated Approach and the Role of Estimate”, Greece, Idea Group Publishing, 2003, pp. 261-285,ch 12 .
7. Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan designed—Web

Usage Mining: Discovery and Applications of Usage Patterns from Web Data], 2000.