# Web Personalisation based on user interaction

## Web Personalisation

Sumit Sakarkar
SCET
MIT Academy Of Engineering
Pune, India
sasakarkar@mitaoe.ac.in

Vaibhav Chaudhari
SCET
MIT Academy Of Engineering
Pune, India
vrchaudhari@mitaoe.ac.in

Tanmay Gaurkar
SCET
MIT Academy Of Engineering
Pune, India
trgaurkar@mitaoe.ac.in

Aditya Veer
SCET
MIT Academy Of Engineering
Pune, India
asveer@mitaoe.ac.in

Mrs. Mayura Kulkarni
SCET
MIT Academy Of Engineering
Pune, India
mukinikar@comp.maepune.ac.in

*Abstract*—The World Wide Web is routinely favored plus a global reference for collecting information. Sadly various web deployed learning systems are providing identical educational resources to different users/profiles. Many work has been done for personalizing systems. In this paper, our objective is to evaluate recommendations of active users/learners by studying her / his history plus utilizing likeness as well as differences between the user preferences together with the content of the learning resources. We have used here concepts like Web scraping, Web usage mining for Web personalization. User profiling and personalizing the web search results are used to evaluate the links which could be recommended to the active users. Web personalization provides users/learner's relevant information and also the information which is facilitated by the likeness of the users/learners. The main motto of personalization is to make a website more flexible (responsive) towards the users/learner's requirements. Web mining which is covering concepts like web content mining as well as web usage mining as well as web structure mining falls under data mining, Personalization also falls under web usage mining. Researchers constantly concentrate toward fulfilling the users/learner's wants, as all users/learners have different backgrounds plus different goals while browsing the web. Search engines should be up to date (aware) with user/learner likeness and preferences so user/learner likeness and preferences should be gathered and used, For this, the concepts like Collaborative Filtering as well as Document Based as well as Concept based profiling can be used.

*Keywords—World Wide Web, Web Personalization, Web server logs, Data Preprocessing, Pattern discovery ,Web scraping, Web usage mining,* [1] *User Profiling* [2]*.*

## I. INTRODUCTION

In the modern era, users demand personalized experiences. They required information that is relevant, individualized, and provided them with an accurate recommendation for the new content and new products based on their previous activities. People usually spend hours searching for relevant products or information over the internet. The motive of this paper is to deliver a personalized experience to the active user while searching for data. Users, when searched data will inform about its relevance as feedback and according to a customized search, will be provided to the users To provide a personalized experience to the user while searching for data. Users, when searched data will inform about its relevance as feedback and according to a customized search, will be provided to the users.

The first step is to collect data from users which may then be divided into either personal data or behavioural data. The personal data includes age, gender and demographic details whereas other one includes usage, click stream and time [3].

Another branch of this could be automatic recommendation , for e.g. Amazon's recommendations and Mystrand.com Music recommendation  and also e-learning smart recommender [4] [5]. Multiple strategies to automate personalization can be

detected in the literature that include content-based . item based, collaborative or rule based filtering and as well as web usage mining . Today, RS (Recommender system) allows the users to get transferred towards customers, rising cross-selling and building loyalty, this being the reason [6] [7].

We are going to implement a personalization system in which users will browngines to give them relevant information as well as accurate information is known as web personalization. Identifying as well as retrieving as well the content according to their requirement then from the searched information they will provide feedback for the relevance of that where the data will be processed through various stages such as editorial planning, content reusing ,navigation and content hierarchy.

The action or the process of collecting and saving the information of users/learners while communicating and interacting with search e as assembling are the steps for successful web personalization. Initially, we have to gather all obtainable data. The data is differentiated into two types that are personal data as well as demographics ( behavioural ) data. In the second step, the search engine creates a users/learners profile plus uses intelligent algorithms to examine and extract the information and find the relation between web pages as well as users/learners' likings. In third and the last step, the search engine will provide users/learners with accurate as well as relevant information

Due to the massive increase in data, there is a significant broadening in the search engine and web based applications as it involves recognizing similar patterns in the data, Web mining comes under the technique of mining data from the World Wide Web. Web usage mining refers to the action of withdrawing functional and helpful information coming from search engines depending upon users browsing patterns. It helps in finding patterns from weblogs. Pre-processing, as well as pattern discovery as well as Pattern analysis, are the stages for implementing Web usage mining

## II.    OBJECTIVES

The main objective of this project is to provide personalized experience to the user while searching data. Users when searched data will inform about its relevance as a feedback and according a customized search will be provided to the users.Also we will try to find useful and relevant information based on the keyword typed and the user's interest which is found out using different methods of personalization of information and algorithms. We'll try to save a whole bunch of time working over the internet and provide more Relevant data Recommendations.

## III.    LITERATURE SURVEY

A paper was published by M. Rami Ghorab, Dong Zhou, Alexander O'Connor, Vincent Wade Published online on 13 May 2012 at the Springer Science Business which focused on Personalized Information Retrieval survey and its classification [8]. It had a mechanism to learn about their users' search interests by either explicitly supplying this information implicitly gathering this information in an unobtrusive manner from the users' search history.. In order to provide a personalised service,this system maintains information about the users and the history of their interactions with the system.   This paper had a three-step approach including the information gathering which included how and which information to be gathered from the user and his usage behavior. The next step included representation of the gathered information either through vector-based models or Semantic network-based Models where user's interests are maintained in a network structure of terms and its related terms.User models here are represented using a semantic network structure. In this case the model is made up of nodes and associated nodes that capture terms and their semantically-related or co-occurring terms respectively.The next step would be implementation of the query adaption by the use of various algorithmic approaches like global analysis, Explicit relevance feedback, interactive query expansion.

Another paper written by Magdalini Erinaki in the year 2003 focused on Web mining for Web personalization [9]. This paper includes personalization techniques such as Content based filtering-recommendation based on individual past ratings and preferences and the Rule based filtering- based on some predefined set of rules. The overall process of usage-based Web personalization consists of five modules, which correspond to each step of the process. These are as follows.

❏    User profiling
❏    Log analysis and Web usage mining
❏    Content management
❏    Web site publishing
❏    Information acquisition and searching

It merely records the addresses of pages requested by its user thus highlighting interesting hyperlinks without involving the user .The main advantage that they achieved was that the search is user specific and the process is quite easier but Both rule based and content based techniques do not provide large scope and hence limiting benefits.

A study was done on another paper published in IEEE in the year 2008 which enlightened Profiling [10]. They worked with Implicit data gathering and personalization focusing on Memory based technique which involves saving past rating for each user for different items and maintaining pairwise similarity between users and also the Model based technique was used where Models are built on past searches and prediction models are generated based on behavioral data collection and Real time recommendation score generation. But the major problem that they found here was that Data Sparsity and inefficient user group categorization can lead to unreliable results.

A further study was done on a paper published in IEEE 2011 written by Mandeep Pannu which focused on Explicit user profiles in web Personalization [11]. This paper includes personalization techniques such as Vector based modelling which involves creating preferences vectors for the users and comparing them with the same vector model created for web data and then comparing them. Also, they used the Probability modelling which worked for estimating the probability of relevance to the ranking document for a query. The main advantage of this paper was that this Allows ranking documents according to their possible relevance but Long documents are poorly represented because they have poor similarity values. This implemented system allows users to create, save, remove and update profiles based on their changing information needs and preferences thus providing a personalised experience. Preliminary experimental results indicate that the system produces more user-interest based results based on explicit user profiles when compared to traditional search methods which provided the same data for a similar search to every user.

A similar study was done and it's proposed framework composed of two modules: an off-line module which pre-processes data to build user and content profiles, and an on-line module which uses these models on-the-fly to recognize user goals and predict a recommendation list to provide more relevant data to users. They applied data mining techniques to build user profiles, where the prediction of the user model is accomplished not using explicit user interaction, but rather implicit information collected from all past usage sessions which is also known as usage behaviour. The input data for this first step consists mainly of Web server access log files. In order to extract useful information from log files and build user profiles, they applied Web Usage Mining techniques.It the application of data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of Web-based applications.

## IV. THEORY AND RELEVANCE

### Web Scraping

This is an automatic version of gathering well-structured data from web pages. It is also known as web data extraction. Web scraping a web page implies fetching it, and extracting data from it. Fetching is actually the downloading of a page. Hence, web crawling is a major component of web scraping, to fetch pages for further processing. Once fetched, then extraction can take place [12]. Web Crawling actually consists of a crawler that automatically searches and explores content on the web. Crawlers are fundamentally programmed for repetitive steps and actions such that web browsing and searching gets automated. Some websites contain a massive amount of invaluable data. It's a very tedious job to manually extract this huge amount of data. Web Scraping extracts this huge amount of data and then exports it to a format that is more convenient for the user ( For Eg. Excel, pdf etc).

### Web Usage mining

Web usage mining is a subcategory of data mining which is used for recognizing usage patterns and trends from web data. This is a combination of different applied techniques, to unravel informative usage trends from data taken over the internet. Web usage mining understands the requirements and needs of web based applications and It is then analyzed, understood and is then fed to web based applications. This type of data captures the identity or emergence of Web users along with their behavior of how they browse and browsing history, at browser level as well as website level [13].

### Web personalization

Web Personalisation or Personalisation of information is a web mining technique which helps the customers to find relevant information to the searches they make on the internet. The main focus of this project is on providing the users the provision of bettering their web searches and personalizing according to their experience and also help them in personalizing the information with some external help such as recommendation system, search feedback, pages indexing etc.

After studying and analyzing many paper and articles on web, we found that the main problem in the system is :

A.    Optimized Landing Pages
B.    More Relevant Content Recommendations
C.    Increase in User Engagement on the Website

D.      Unsatisfactory Search Results
E.      Inefficient methods of getting useful information
F.      Reduce the time required for finding useful results

### *Beautiful Soap and Parsehub*

Beautiful Soap is an open source python library used for scraping data into our convenient format. We have used beautiful soap along with Parsehub which makes it easier to modify, navigate and search the generated parse tree. Parsehub is used in the program development to access various features such as scraping the data from different websites and store the scraped data in a machine understandable format i.e. in a proper database.

### *Proposed System Design*
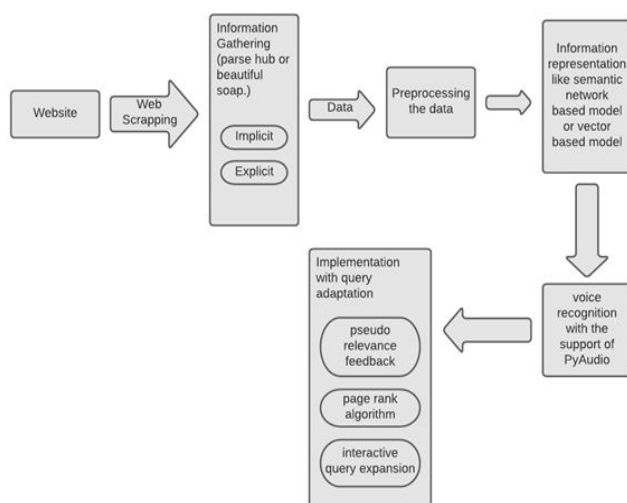


## V.    SYSTEM FEATURES

### Planning :

An editorial style-guide which includes voice tone, phrasing and content length etc .A content gap analysis to ensure that all the relevant content is available for each customer segment.

### Content Reuse :

Optimized content formatting is ensured across devices. A review of tokens which will be used only in content to permit for changing data which is intended to be changed counting on visitor context.

### Content Structure , Taxonomy and Metadata :

A content Model document , included within the User Experience that describes the structure of cotet as well as taxonomy. This contains proper documentation for checking that all types of tags such as meta tag, headings tag, title tags as well as rich snippets are used based on correct ways

## VI.    METHODOLOGY

The major functions that product must perform are

### Scraping of data

Web scraping is the procedure of gathering structured web data in an automated method [14]. It is also known as web data extraction. Web scraping extracts the primary HTML code and, with it, data reserved in the database.

### Analysing and preprocessing the data

We will be analysing the extracted (scraped) data and will be applying all the pre-processing steps needed to make the data less noisy and more processable.

### Classification to show specific results to the users

The searched product will be categorised and classified according to the users need and accordingly results will be displayed

### Applying Algorithm

Suitable algorithms will be used in order to enhance the system performance.

### Speech Recognition

We are adding voice assistance to it, in order to make it more user friendly. The input (query) for the speech recognition can be managed as per the user's requirements [15].

## VII.    Work Done

We have successfully scrapped the data from different sources and completed all the pre-processing steps of data and the next step is to apply the suitable algorithm for web personalization. For the start of the venture, data from CoronaVirus counter website is scraped and stored as a dataset input to the python program which works as the audio user interface for the user where data on the website can be shared with the user in form

of sound. In the next step, data from e-commerce websites is scraped and made available in a similar way.The further steps include personalisation of the data as stated in future Scope.

## VIII. FUTURE SCOPE

Working with the internet provides a lot of security breaches and a whole bunch of time is wasted searching for the relevant information.The future scope of this paper is to design a framework to provide a complete privacy of the users so that they can work without any fear of working in a personalized web environment [16]. Also, we will try to make the algorithm more efficient and the overall time complexity of the whole program should be minimized.

## IX. CONCLUSION

Product recommendation or personalization is the procedure of generating customized experiences for users to a particular website or content. Instead of providing a single, wide exposure, Web recommendation would allow organizations to present users with distinctive and unique experiences customized to their needs and expectations and saving a lot of time. We have successfully scrapped the data from different sources and completed all the preprocessing steps of data and the next part is to apply the suitable algorithm for web personalization which will give the most efficient results.

But, it is impossible to completely personalize a particular website for a particular user. Web Personalization creates consistency among the website users and every user gets a personalized experience of the website but at the same time we cannot get it right everytime that's what's going on in the mind of the user. Web Personalization comes with great customer service and satisfaction although it comes at its own cost for the users.

### REFERENCES

[1] F. M. f. Pier luca lanzi, "Recent Developments in Web Usage Mining Research," *5th International Conference,* pp. 3-5, 2003.

[2] S. G. D. M. Suraj Kumar Kanoje, "User Profiling for Recommendation System," March, 2015.

[3] A. f. M. R. ,. M. N. Kardan, "Proposing an architecture for learner modelling based on web usage mining in e-learning environment," *The 5th Data Mining Conference,* pp. 14 - 15, 2011.

[4] B. Mobasher, "Data Mining for web Personalisation," *Springer verlag,* 2006.

[5] O. R. Zaiane, "Building a Recommender Agent for E-Learning System," *7th International Conference on Computers in Education,* pp. 3 - 6, 2002.

[6] J.D.P.C. Velaquez, "Building a knowledge base for implementing a web based computerised recommendation system," *International journal on Artificial Intelligence Tools,* vol. 16, no. 5, Oct. 2007.

[7] A. Roshanzamir, "Web Personalisation Implications and challenges," *ISCET.*

[8] R. Ghorab, "Personalised Information Retrieval: survey and classification," *Springer Science+Business Media B.V. 2012,* 2012.

[9] M. V. Magnalini Eirinaki, "Web Mining for Web Personalisation," *ACM Transactions on Internet Technology,* pp. 1-27, 2003.

[10] "Personalisation of Web Search results Based on User Profiling," *1st International Conference on Emerging Trends in Engineering and Technology,* 2008.

[11] m. pannu, "Explicit user profiles in web search personalisation," *Proceedings of the 2011 15th International Conference on CSCWD,* 2011.

[12] R.B.S.K.R.S.M. Bela Jogalekar, "Search Engine Optimisation using Unsupervised learning," *5th International Conference on Computing, Communication, Control and Automation,* 2019.

[13] P. Chavan, "Web Usage Mining," *ISCET.*

[14] "Explicit User Profiles in Web Search Personalisation," *15th International Conference on Computer Supported Cooperative Work in Design,* 2011.

[15] S.K.E.M.A. Anusuya, "Speech recognition by machine : A Review," *International Journal of Computer Science and Information Security,* Sept. 2009.

[16] S. Sayed, "A Survey of Web Page Personalization in Web Search Engine," *Web Page Personalization,* vol. 2, no. 1, p. 4, 1 january 2016.

[17] J. Blom, "Personalization - A Taxonomy," *CHI: Conference on Human Factors in Computing Systems,* p. 314, 2000.

[18] a. deore, k. shah and u. pawar, *International Research Journal of Engineering and Technology,* vol. 7, no. 9, 2020.

[19] O. b. Postma, "Personalisation in practice," *Springer ( palgrave ),* pp. 137 - 142, 2002.

[20] O. Nasraoui, "World Wide Web Personalisation," *Invited Chapter in Encyclopedia of data Mining and Data warehousing,* 2005.

[21] S. A. G. Aanand V. Saurkar, "An Overview on Web Scraping Techniques and Tools," *International Journal on Future Revolution in Computer Science and Communication Engineering,* vol. 4, no. 4.