

**GLOBAL DYNAMIC JOURNAL OF ENGINEERING, PHYSICAL AND BIOLOGICAL SCIENCES,  
HUMANITIES, EDUCATION AND AGRICULTURE JOHANNESBURG, SOUTH AFRICA**

**Volume 1, Number 2, 2022, ISSN: 2955-053X Indexed in Google Scholar**

**Email: [dynamicjou@gmail.com](mailto:dynamicjou@gmail.com)**

# **GDJEPHEA**

**GLOBAL DYNAMIC JOURNAL OF ENGINEERING, PHYSICAL AND BIOLOGICAL SCIENCES  
HUMANITIES, EDUCATION AND AGRICULTURE  
JOHANNESBURG, SOUTH AFRICA**

**PUBLISHED BY GLOBAL DYNAMIC JOURNAL OF ENGINEERING, PHYSICAL AND BIOLOGICAL SCIENCES  
HUMANITIES, EDUCATION AND AGRICULTURE  
JOHANNESBURG, SOUTH AFRICA**

**ISSN: 2955-053X**

## AN IMPROVED PREFIX-BASED ALGORITHM FOR WEBLOG USERS RECOMMENDATION SYSTEM

E.F. Jumbo <sup>1</sup> & C. J. Nwagbara <sup>2</sup>

Department of Computer Science,  
University of Port Harcourt, Port Harcourt, Nigeria.

### **Abstract:**

It is becoming rapidly difficult to get log-information test from Web as similar to related need, log-information is overloaded in the Internet and getting the needed log-information from Internet is become hard task and as such, there is a need to use a good mechanism to filter information based on user need. This paper proposes an improved prefix-based algorithm for weblog users' recommendation system in filtering weblog information using prefix & suffix search to obtain relevant information need for the weblog user. The system uses Structure System Analysis and Design Methodology (SSADM), PHP Programming language and MYSQL database in developing the system. The result prefix and suffix search enable the Weblog system visitors, easy navigation to the web pages, quickly reaching their destination and to obtain relevant information.

**Keywords:** Algorithm, Prefix-Base, Recommendation, Weblog.

---

**Background:** In background of Web-search, it is hard to actually identify anticipation and intends of user. Web-based recommendation systems help website visitors or users to easily navigate the web pages fast, getting to their terminus and obtain crucial details. There are two approaches for developing recommendation procedure systems: Content-based filtering procedure, two Collaboration-based filtering techniques, some situation combination or hybrid of both techniques are preferred by researchers. Shiva et al., (2011) showed that Content-based filtering procedure is design that is dependent on content gathered from target items. In this technique, web pages are commended for user completely based on profile designed by analyzing content of things which such user rated in formal. The user profile are constructed through investigating responses to questionnaire, item ratings, or user steering information to infer preferences of user's interests. They therefore concluded that using content-based technique recommendations are done mostly from past experience of web visitor. In collaboration filtering technique, web pages are commended to particular user when other similar users equally prefers such web pages (Riecken, 2010). This definition of 'similarity' between wed-users is dependent on applications that use web commendation system. For instance, it is sometimes defined as users with similar web-pages rating or web-sites or users with associated navigation features. A collaboration filtering technique collects entire information concerning activities on web site from web-servers and calculates similarity among users. The demerit of this technique is that users do not present or fill their rating often or properly for website or web-pages even when it is designed to helped them in future search. Several researchers employed web-using mining as commendation tool. Web-usage mining is divided into two phases namely Data Preparation and Pattern Discovering. During preparation stage, data are sourced from web-server that stores or files user session details then make them appropriate for applying.

### 1. Aim and Objectives of the study

The aim of this work is to develop an improved prefix-based algorithm or pathway for web-user commendation system and their specific objectives are to:

1. Develop an improved prefix-based algorithm for weblog users commendation system
2. Implement this model using PHP programming language with help from FP-Tree algorithm
3. Compare this developed model with the existing search algorithm.

### 1.2 Significance of the Study

The smart algorithm for web-user commendation system is helpful to commendation engine because it would trigger influx of traffic to their site. It would accomplish such with

modified email messages. By analyzing present site used by customer and their formal browsing history, commendation engine could deliver crucial product commendations when next or as they are presenting shopping. This information are sourced in real-time so software could react when their shopping features alters. Shoppers are engaged or involve in site more especially when personal product commendations are made. They would delve deep into product without performing several searches. Converting mere window shoppers to real customers take special dealing and handling. Personal connection from commendation engine reveals to your customers they are worthwhile on personal basis, which in turn triggers or engenders loyalty. It equally elevates average order worth when commendation engine are employed to show personal options to users and customers. Advance metrics and documenting would show effectiveness in such venture definitively.

## **2.1 Related Work**

Web commendation procedures forecast details requirement customers or users and present them with commendations to facilitate their searching. Several methods like web-content mining and web-structure mining are designed to support web endorsement system. However, these techniques encounters some main setbacks where most customers and users anonymously surf web-sites through proxy with their identities hidden and hard to reach. Web-user mining (Baoyao et al., 2010) is among the key techniques to review users' surfing or searching patterns and web-resources use like web-server access, proxy-server, browser-logs, customers profiles, registered data, customers sessions, customers queries, and others to understand customers for better service on Web-based applications.

Recommending systems generally produce or generate detail of recommendations in any of these ways one, through collaboration filtering or content-based filtering equally called personal-based technique (Osein et al., 2012). Collaboration filtering technique develops model using customers formal features like thing bought previously or things selected or number based ratings that was given and similar decisions which was made by other similar user, it is then employed to forecast or predict items which involved customer or shopper may have interest. Content-based filtering technique uses several of discrete features of item to endorse other items having similar features (Mooney and Roy, 1999).

The disparity between collaboration and content-oriented filtering are shown through comparing two common music recommending procedures – Last-search and Pandora Radio.

- i. Last in creating "station" for endorsed songs by noticing bands and personal tracks which user listens to regularly and comparing that with listening features from other users. Last-search plays tracks which are seen in user library but are usually played by other customers with similar flare. As such technique leverages features of users, it is example of collaboration filtering procedures.
- ii. Pandora utilizes features of song or their artist to seed stations which plays music having same or similar features and feedback from such user are used in refining results of such station, deemphasizing some features when such user "dislikes" certain song and emphasize other features when such user "likes" selected song. This is example of content-oriented technique.

These two forms of system have their peculiar strengths and shortcomings. In these instances made above, Last-search needs huge stock of information on user before making informed endorsement or commendations. This is example of cold start issue and usually noticed in collaboration-oriented systems (Rubens et al., 2016). Recommender procedures are useful options for searching algorithms because they aid users to uncover items which they may not have uncovered. Note that recommending systems are usually implemented with help of search engines.

### 2.1.1 Approaches used in Recommender System

#### a. Collaboration filtering

One technique to recommending systems designers which possess wider use is collaboration-oriented filtering (Sarwar, et al, 2000) These techniques are based on sourcing and analyzing huge amount of detail on user features, function or preference and forecasting what customers might want based on similarity from other customers. A key benefits of collaboration filtering technique is that it don't depend on engine analyzable content and it is able to correctly recommend complicated items like movies without "understanding" of item themselves.

Collaboration-based filtering procedure are grouped as memory and model oriented collaboration filtering procedure. Well-oriented example for memory-oriented procedure is user-oriented technique and model-oriented technique is Kernel-Map Recommender (Ghazanfar et al., 2012).

#### b. Content-based filtering

Another technique employed when developing recommender procedure is **content-oriented filtering**. This method is based on describing item and profile of preferences made by users (Aggarwal, 2016). In this system, key-words are employed in describing items and customers profile is designed to show kind of item such customer likes, meaning that involved algorithms attempt to recommend product which are same or similar to that from other customer likes in past surfing or shopping. In particular, several candidate products are compared to products rated previously by customers and best-matched product are recommended. This technique are rooted in details filtering searching.

To nonconcrete features of product in this system, products presentation procedure is applied. A commonly utilize procedure is tf-idf representation equally called vector-spaced representation.

### 2.2 Weblog

A **blog** which is short form of term "**weblog**" is discussion website that is usually published on global wed-site that consist of discrete, informal diary, style text entries. Posts are displayed and viewed in opposite sequential order such that most current post occur first, on top of their web-page. Not until 2009, blogs are mostly dwelling platform for single individual according to Blood, (2000). Occasionally small group, and covered single details or topic. In 2010, "multi-author blogs" emerged and featured writing from several authors and professionally reviewed. MAB from papers and other media, universities, thinking tanks, support partners and similar institutes account for increasing blog traffic

quantity. Rise of Twitter and "microblogging" systems aided to incorporate MAB and one-author blogs news media. Blog are also used as verb which means to maintain or introduce content to blog.

**Zhou and Hui (2015)** presented CS-Mine procedure that is effective compare to WAP and PLWAP mining by removing entirely need for recreating middle condition in WAP-trees. However, it equally needed to create initial WAP-tree direct from web-access database firstly. The "Pattern Growth based algorithms" utilized compacted data model known as WAP tree that files web-access patterns, and effective technique for uncovering common patterns. The already existing web commendation procedure employed condition sequence mining procedure where WAP tree are created initially. Though WAP tree is flattened data assembly, time needed for WAP tree generation increases appreciably when mean length of web-access sequence elevates. The WAP tree assembly uses optimum distribution of simple prefix paths in tree creation. When web-access possess little unique prefix paths, therefore WAP tree created are massive. To uncover "Prefix Sequence Bases" of every common token, the entire WAP tree are negotiated every time in top-bottom technique. Massive amount of common tokens elevates WAP-tree traversal duration.

### 3. Material And Methods.

**3.1 Analysis:** Analysis is very important as a phase in system development life cycle where real data are collected, understood and processed to identify problems and recommend solutions to enhance functionality of involved system. This is done to produce new idea that satisfies the present or current needs of user and allow for further and future improvement.

#### 3.1.1 Analysis of the Existing System

The existing system is based on Zhou and Hui (2015) proposed a CS-Mine algorithm which is efficient than both WAP-Mine and PLWAP-Mine by removing entirely need for recreating middle condition in WAP-trees. However, it equally needed to create initial WAP-tree direct from web-access database firstly. The "Pattern Growth based algorithms" utilized compacted data model known as WAP tree that files web-access patterns, and effective technique for uncovering common patterns. The already existing web commendation procedure employed condition sequence mining procedure where WAP tree are created initially. Though WAP tree is flattened data assembly, time needed for WAP tree generation increases appreciably when mean length of web-access sequence elevates. The WAP tree assembly uses optimum distribution of simple prefix paths in tree creation. When web-access possess little unique prefix paths, therefore WAP tree created are massive. To uncover "Prefix Sequence Bases" of every common token, the entire WAP tree are negotiated every time in top-bottom technique. Massive amount of common tokens elevates WAP-tree traversal duration.. Figure 3.1 illustrate the CSB-Mine architecture.



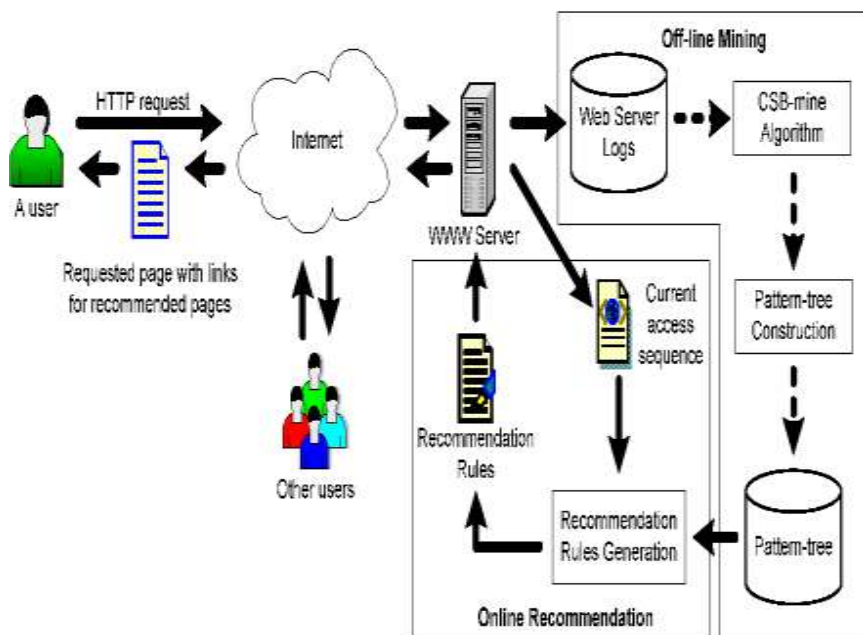


Figure 3.1: Architecture of the Existing System (Zhou and Hui, 2015)

### 3.2 The Proposed System

The proposed system will improve on Zhou and Hui (2015) by introducing a smart procedure which would eliminate the need for creation of an initial WAP-tree. The MCS-Mine procedure, which considers suffix sequencing bases for every common token instead of prefix sequencing, is illustrated in Figure 3.3. This figure shows the proposed system structure or assembly and elaborates on how adapted CS-Mine works. The proposed procedure offers a list of recommenders' web-pages to active users without the need to access history or a minimal amount of visited pages. A web user can be any one who is accessing an online web site. A web user could be a naive user or a trained one. When a web user accesses a particular web-page, an entry is made in the web-log file by the web-server. The web-log file comprises of information relating to user web access, such as client IP address, demanded time, demanded URL, user-ID, HTTP status code, etc. The raw web-log file contains many unwanted information for the mining process. Therefore, web log files are cleaned, which is conducted via web log file preprocessing.

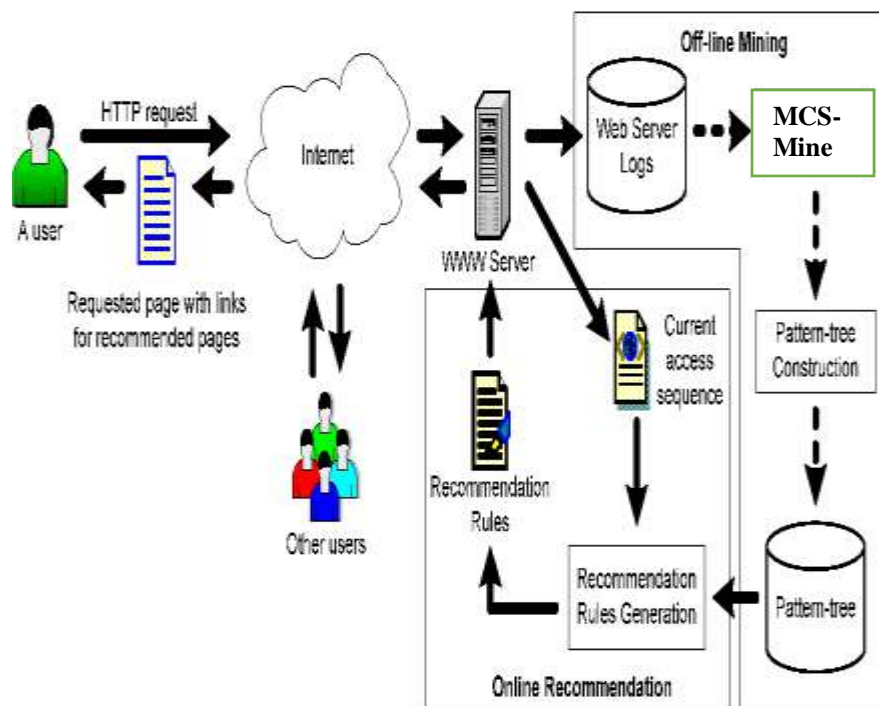


Figure 3.3 Structure of Proposed System

The input to this block is raw web log file and output is “web access sequence database” (WASD). After pre-processing MCS-Mine algorithm is applied for finding sequencing patterns which are stored in text file and are recommended list. The figure 3.4 shows Web Log Preprocessing in detail.

### 3.2.1 Algorithm of Proposed System (MCS-Mine):

Step 1 Read WASD and for every common token, find suffix sequencing base (SSB).  
 Step 2 Use Construct EQ to construct event queues for SSB (token).  
 Step 3 Use Test SSB to test single sequencing for final-SSB (token).  
 If test is successful,  
 Create frequent sequence FS = token+SingleSeq and insert into WAP List.  
 Else,  
 For each token t in Header Table of final-SSB, use Construct Sub SSB to construct final-SSB (token+t).  
 Set token = token+t and recursively mine final-SSB (token) from step 2.  
 Step 4 Return WAP List.

## 4. Results

From all indications of the resulting output, the results from the implementation of the software is précised on the security of the warehouse Data presented and variables were studied, a well predictable set of data were used and graphics representing each state of instances. The table 4.1 shows the parameters for experiment 1 in order to discover the beauty of the proposed system. If a user logs in



successfully, he/she will have to search and read the document searched for if it exists in the Database.

### Experiment 1

From the experiment in table 1, the application is designed to help or enable users enter the keyword that will aid the searching of document(s) in the proposed system database and if the document is available, the user can view the document while the application still running. But if the document is unavailable for any reason, it means that the users cannot view it.

Then the user has to view related works attached to the database to avoid searching for non-existing document. The value 1 as shown in the table above indicates that the user cannot use keywords simultaneously on the search box.

The table 4.1 shows the parameters for experiment 1.

Parameter	Value
Keyword	1
Search document	1
Available	1
Unavailable	0

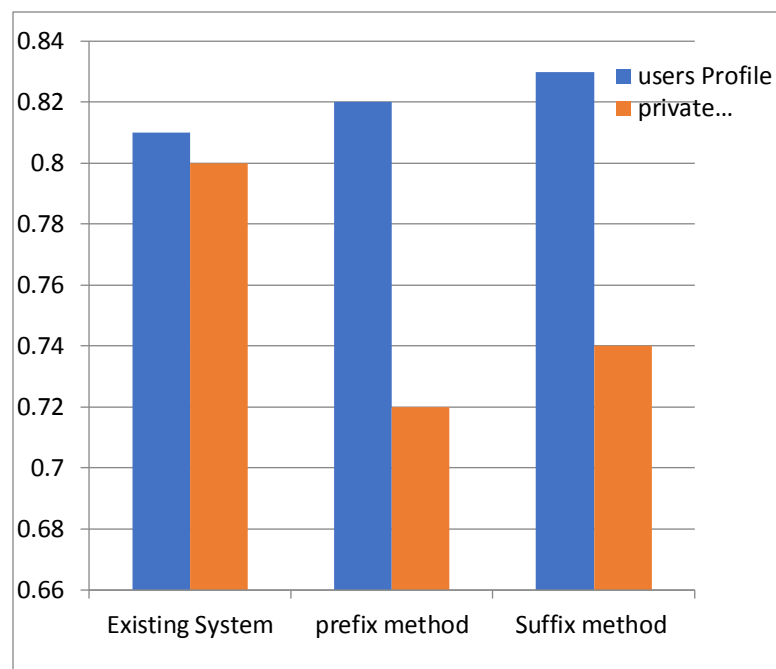
### Experiment 2

Table 4.2 shows experiment 2 where we investigate our program speed and efficiency in searching for document like (improved system using hybrid method). This research will focus on using web-usage mining to obtain customers profile and group customers sessions using “hierarchical unsupervised niche clustering method”. The intelligent algorithm will be able to extract the users profile that has ever been register on such platform must be noted for other friends to make use of it. Other result are shown in figure 4.4 to 4.9.

Table 4.2: Experiment 2

	Existing system	Proposed system

Users profile (keywords)	0.94	0.84
Private information	0.64	0.7
Prefix	0.82	0.72
Suffix	0.83	0.74



**Fig 4.1 Graph of Suffix and Prefix System.**

## **5. Conclusion**

In conclusion, this thesis uses hybrid technique to provide personalized recommender of blogs by fusing different recommender techniques and text mining. The system uses text mining to classify blogs to suitable group and collaboration filtering to group users with similar preferences and demographic profiles. It creates ranked lists of suggestions by computing final scores from different recommender techniques and discusses difficulties

tied to designing recommender systems. The presented system tries to minimize sparsity and slow-start challenges and aims at improving quality and accuracy of personalized blog recommender.

## References

- Sarwar, B., Karypis, G., Konstan, J., and Reid, J. (2001). Item-based collaborative filtering algorithms. GroupLens research group, Army HPC research center, University of Minnesota, Minneapolis, 1-11.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2000. Analysis of recommendation algorithms for e-commerce. Proceedings of ACM Ecommerce Conference. 158–167.
- Robin Burke, [Hybrid Web Recommender Systems Archived](#) 2014-09-12 at the [Wayback Machine](#), pp. 377-408, The Adaptive Web, Peter Brusilovsky, Alfred Kobsa, Wolfgang Nejdl (Ed.), Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, Lecture Notes in Computer Science, Vol. 4321, May 2007, 978-3-540-72078-2.
- Eirinaki, M., Vazirgiannis, M., and Varlamis, I. 2003. SEWeP: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process. SIGKDD '03, August 24–27. Washington, DC, USA. Copyright 2003 ACM 1-58113-737-0/03/0008.
- Ekstrand, M., Riedl, J., and Konstan, J. (2010). Collaborative Filtering Recommender Systems. Foundations and Trends in Human-Computer Interaction, 81-173.
- Ekstrand, Michael D.; Ludwig, Michael; Konstan, J.; and Riedl, John T. (2011). Rethinking the Recommender Research Ecosystem: Reproducibility, Openness, and LensKit. Proceedings of the Fifth ACM Conference on Recommender Systems. RecSys '11. New York, NY, USA: ACM. pp. 133–140.
- Ghazanfar, M.A.; Prügel-Bennett, A.; Szedmak, S. (2012). "Kernel-Mapping Recommender system algorithms". Information Sciences. **208**: 81–104.
- Gomez-Urbe, Carlos A.; Hunt, Neil (28 December 2015). "The Netflix Recommender System". ACM Transactions on Management Information Systems. **6** (4): 1–19. [doi:10.1145/2843948](https://doi.org/10.1145/2843948).
- Gretzel, U., and Fesenmaier, D. (2003). Implementing a Knowledge-based Marketing information System: The Illinois Tourism Network. Journal of Information Technology and Tourism, 6(4), 245–255.
- Herlocker, J. L.; Konstan, J. A.; Terveen, L. G.; and Riedl, J. T. (January 2004). "Evaluating collaborative filtering recommender systems". ACM Trans. Inf. Syst. **22** (1): 5–53.
- Hosein J; A.T.H. Sim and R. Saadatdoost (2012): [A Naïve Recommendation Model for Large Databases](#), International Journal of Information and Education Technology.