



Comparative Analysis Of Suffix And Hybrid Algorithm For Weblogger System

E.F. Jumbo & Dr. F. Egbono

**Department of Computer Science. & Department of Information Technology.
University of Port Harcourt, Choba, Nigeria.**

ABSTRACT

Search Compartmentalization is becoming increasingly difficult to access information from Web as information related to a particular need, information is overlocking in the Internet and getting the needing information from Internet is intricate task by which there is a need to use a good mechanism to filtering the information based on user fast need. In this project, a comparative analysis of a suffix and hybrid algorithm for weblogger system is proposed for use, in getting data weblog information by using suffix-search and hybrid-search to obtain a relevant information need for the weblogger users which as a result that improved the existing system implementation on a search algorithm. The system uses an Object Oriented System Analysis and Design Methodology (OOSADM), PHP Programming language and MYSQL for the database implementation and developing the system. This proposes system can help Weblogger system visitors for easy navigation of web pages.

Keywords: Hybrid-search, Suffix-search, Recommendation, Algorithm, Weblogger.

INTRODUCTION

In recent times, scales of Internet is becoming larger, it is now rather difficult to access crucial information from Web. Web recommendation is customarily associated with web-based personalization to conduct needs of buyers or customers. Web-based personalization are task of developing web-based information systems that is appropriate for implication and viewpoint of user. Personalized website noticed their users and collect details or information which they might need based on anticipation in attempt to streamline or modify website contents. Web-based commendation is among the procedures of web-based personalization which is called recommendation of hyperlinks which are integrated to interests and partialities of user. The handlers of these recommended connections or links is carried out either in separate procedure of Web-based page or pop-up window. Web-based commendation systems aid website visitors or user to easily navigation the web pages fast, getting to their terminus and obtain crucial details. There are two approaches for developing commendation procedure systems: one, Content-based filtering procedure, two Collaboration-based filtering technique, some situation combines both technique are preferred by researchers.

2. Related Work: A **blog** which is short form of term "**weblog**" is that is usually published on global web-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).

The introduction and increment of blogs in 1990s accorded with beginning of web-publishing tools which simplified posting of content-based by zero technical customer who don't have enough expertise with HTML or computer program. Formally, knowledge of technology as HTML and FTP was needed to

publish content-based Web and Web customer therefore attended to computer fanatics and hackers. In 2010 most are interactive and Web 2.0 webpages which allow visitors to drop web-site comments, and it is interactive such that differentiates them from static web-pages (Mutum and Wang, 2010). In such sense, blogging is considered as some kind of social network service. Bloggers do not just generate content-based post on blogs, but usually develop social connection with readers and bloggers (Gaudeul, and Peroni, 2010).

2.2 Hybrid recommender systems

Current research showed that hybrid technique, combined collaboration-oriented method and content-oriented technique are more effectual in certain cases. Hybrid technique are executed in various ways, through making content-oriented and collaboration-oriented forecasting separate and combined them, by accumulating content-oriented capacity to collaboration oriented technique or unifying technique into single model (Herlocker et al., 2004). Several research experimentally compare performance of hybrid and complete collaboration oriented technique and showed that hybrid technique provide better recommending than complete technique. These technique is equally employed in overcoming some common issues in recommending systems like slow-cold start and sparsity issues.

2.2.1 Hybrid Filtering System Algorithm

Current research showed that hybrid technique, combined collaboration-oriented method and content-oriented technique are more effectual in certain cases. Hybrid technique are executed in various ways, through making content-oriented and collaboration-oriented forecasting separate and combined them, by accumulating content-oriented capacity to collaboration oriented technique or unifying technique into single model (Herlocker et al., 2004). Several research experimentally compare performance of hybrid and complete collaboration oriented technique and showed that hybrid technique provide better recommending than complete technique. These technique is equally employed in overcoming some common issues in recommending systems like slow-cold start and sparsity issues.

2.3 A Review of Intelligent Systems

Intelligent or Smart systems are automated or computerized system which possess capacity to describe, reason and interpret details which usually result to learning and analyzing data to obtain or extract useful patterns, obtain novel information, and pinpoint technique and characters to sanction action on analysis results (Burkat et. al, 2014).

Intelligent or smart systems are designed in several forms and possess several applications, like processing of datasets, control of drones and robots and others. The concepts utilized in intelligent or smart system design are extracted from various fields which are Artificial Intelligence, Engine Learning, and other fields like psychology, morphology and intellectual science.

In IT, systems are termed or considered as combination of elements or several components that are uniquely related to perform or serve specified purpose. Based on this sense, it could be stated that intelligent or smart systems are composed of atomic component rather it is combination of several unique and related components arranged orderly with main of ensuring workability of entire system as a unit capable of showcasing smartness which entails capacity to reason, perceive, describe and interpret imputed details or data to create useful information.

Hence, intelligent or smart system components are usually artificial intelligence oriented components like expert system, chat robot called Chat-Bot, machine learning apps and other software.

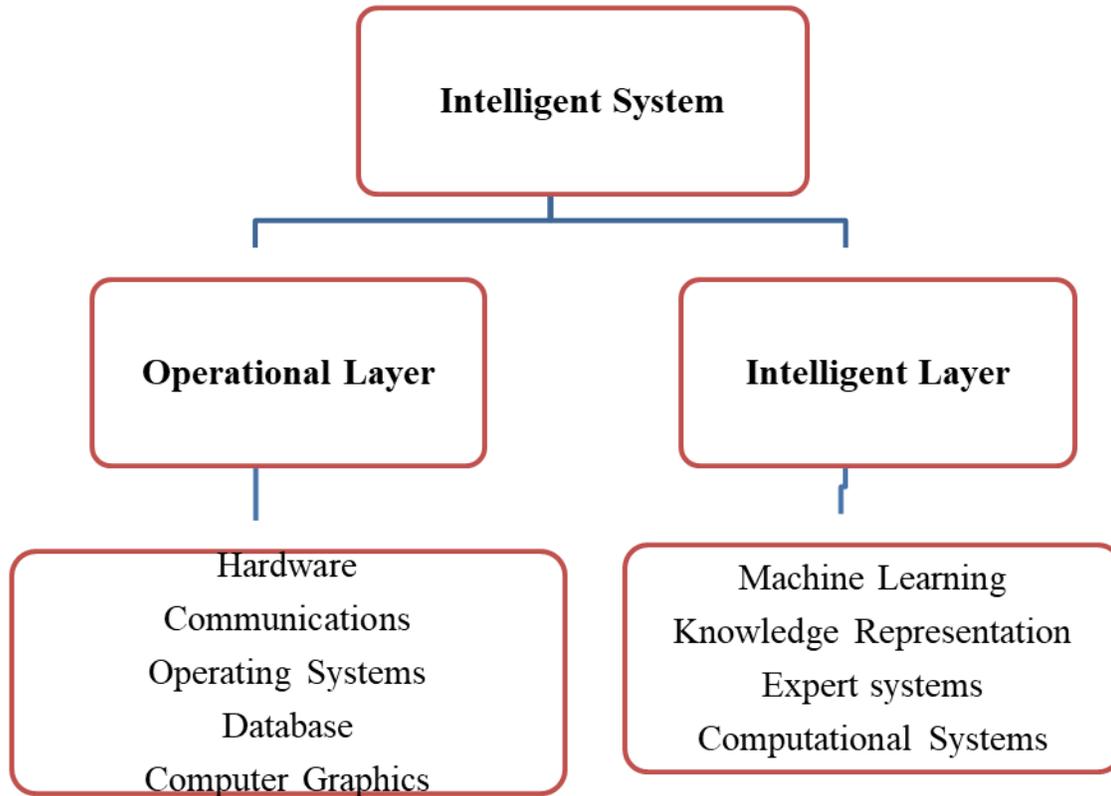


Figure. 2.1 A Hierarchical diagram of an Intelligent System (Dounis, 2010).

It is crucial to state that this software are designed to run on hardware most times like Robotic hardware or directly on private computer hardware system. Hence, these components equally constitutes functional aspect of smart system.

Figure 2.1 show categorized representation of intelligent or smart system with various components which make up such system. These components are separated into two sub-layers which are operation layer and smart layer. Operation layer contains other components which are needed for general operations of smart system. They are hardware by which customers or users sends input into software, interactions that models user’s connection with smart system via physical interaction links energy to operating system; Operating system are underlying software which controls or coordinates entire operations in smart system; also database handles data that are captured by smart system. using these data, they performs crucial operations and conduct numerous queries to assure effectual information retrieval which are needed by smart system to improve learning; Finally, in operation layer of smart system, computer graphics are crucial and collaborating GUI which keeps smart system customers concentrate on supplying needed input and extract output (Dounis, 2010).

2.3 Nearest Neighbor Systems

This is equally called K-Nearest Neighbor or K-NN systems designed using simple and easy machine learning procedure which group input by utilizing their closeness to neighbor. In practice, when closest neighbor procedure was presented with input data that represent example weight or height of specific female. To ascertain gender of specific input, K-NN considers closest k neighbors example k=5 and group this gender of presented input, male. This technique involves simplifying unknown inputs with appreciable success rate

The KNN procedure simply file entire training data, in such case word-based descriptions of clear and implied named product in memory. To group new and unnamed product, this procedure compares it with

entire filed product using similarity operator and ascertain closest neighbor. The group name or numeric code for formal unseen product are obtained from group name of closest neighbors. The similarity operator utilized by closest neighbor procedure depends on kind of data. For organized data, distance metric is commonly employed. When utilizing vector space design, cosine similarity indicator is usually utilized (Salton, 1989).

2.3.1 Features of K-NN Algorithm

a. Between-sample geometric distance

The k-NN classifier is simply based on Euclidean distance existing between test sample and quantified training. Let \mathbf{x}_i represent input having p attributes $(x_{i1}, x_{i2}, \dots, x_{ip})$, n represent number of entire input $(i=1, 2, \dots, n)$ and p entire attributes $(j=1, 2, \dots, p)$. Then Euclidean distance existing between sample \mathbf{x}_i and $\mathbf{x}_l (l=1, 2, \dots, n)$ is given as

$$d(\mathbf{x}_i, \mathbf{x}_l) = \sqrt{((x_{i1}-x_{l1})^2 + (x_{i2}-x_{l2})^2 + \dots + (x_{ip}-x_{lp})^2)} \quad (2.1)$$

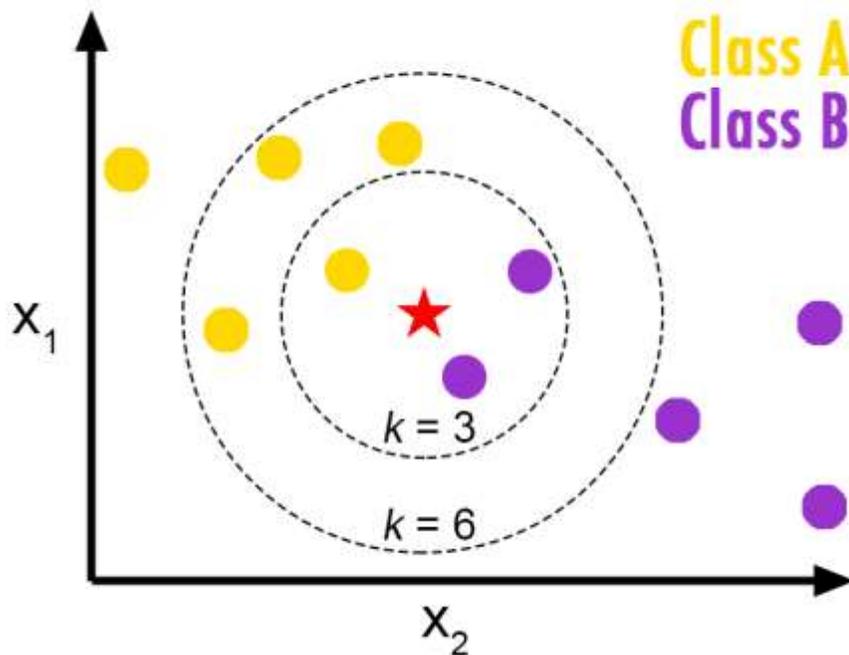


Fig. 2.1 K-Nearest Neighbour Algorithm Concept (Bailey et al., 2016)

Using latter characteristic, K-NN grouping rule is assigned to test sample most category name for k closest training which by practice, k is mostly selected as odd to prevent ties. Then $k = 1$ rule is commonly known as closest neighbor grouping rule.

2.4 Intelligent System Algorithms

Intelligent or Smart systems are automated or computerized system which possess capacity to describe, reason and interpret details which usually result to learning and analyzing data to obtain or extract useful patterns, obtain novel information, and pinpoint technique and characters to sanction action on analysis results (Burkat et. al, 2014). Intelligent or smart systems are designed in several forms and possess several applications, like processing of datasets, control of drones and robots and others.

2.4.1 Content-based System Algorithms

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. To create customer profile used system usually concentrate on two kinds of details: 1. Model of customer's preference. 2. History of customer's connection with recommending system.

3. Material and Methods: 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 in order to improve the functionality of the system. This is done to produce a new idea that satisfies the current needs of the user and allow for future improvement.

3.1.2 Proposed System

The proposed system will as mush improves on Zhou and Hui (2015) by bringing smart processing which will eschew WAP-tree. MCS-Mine algorithm which considers suffix sequencing and CS-mine algorithm which considers Hybrid bases for every common token instead of suffix sequencing. Web-log file authorize comprise prefix and suffix of information relating to user weblog IP address and comparable token is giving to server for usable weblogger. The raw data web-log file contains many unwanted information for mining process. Then web log file are claim to log file preprocessing.

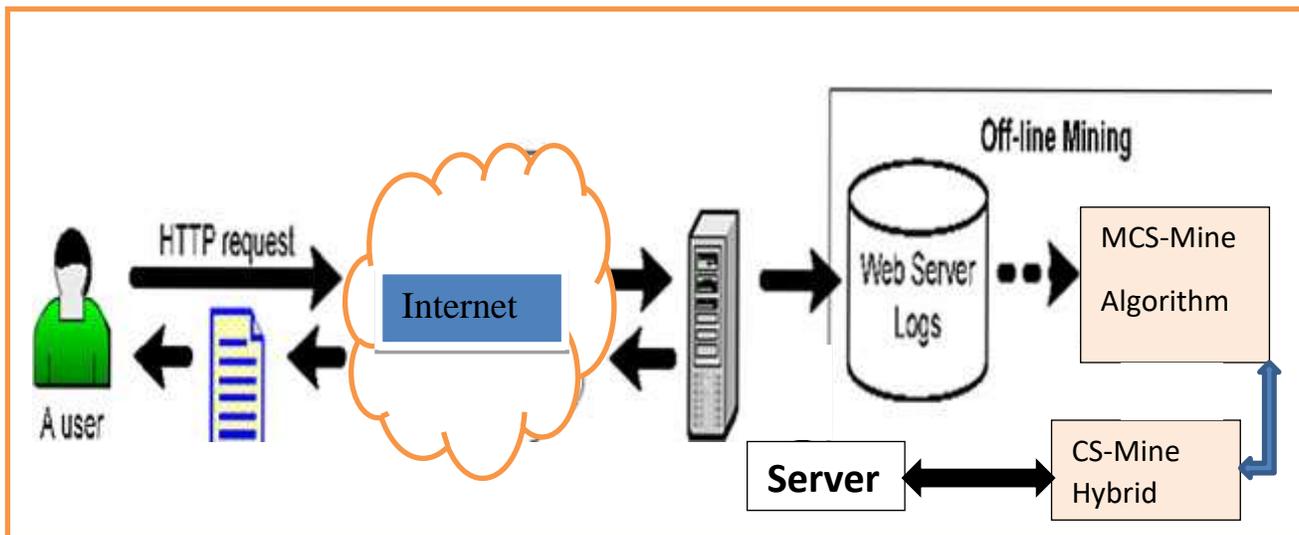


Figure 3.1 Structure of Proposed System

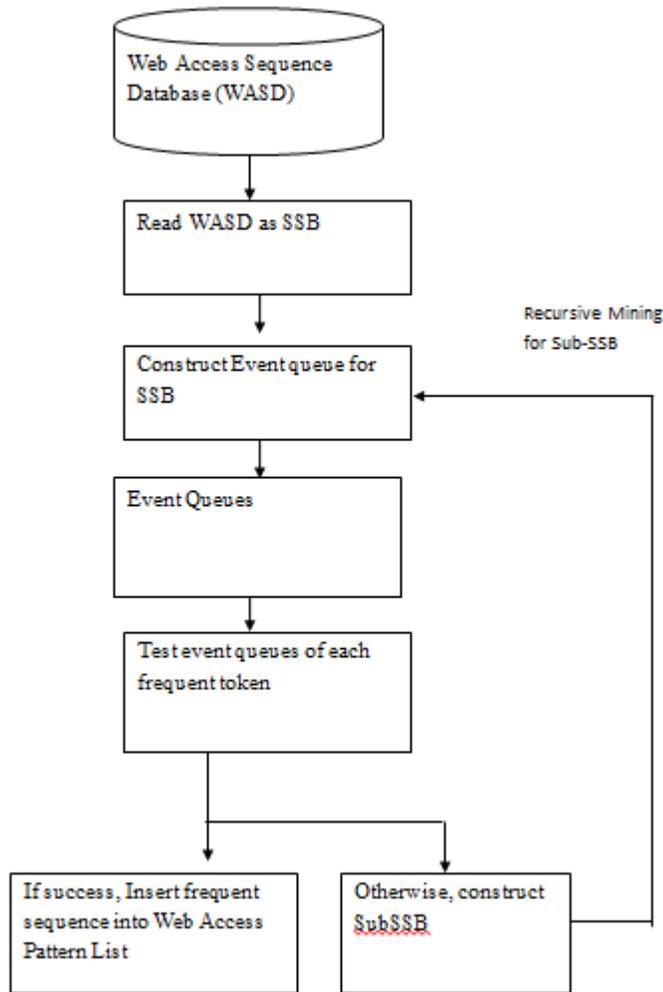


Figure 3.2: MCS-Mine Procedure Data Flow

Procedure and Algorithm Test SSB:

Step 1 Initialize the single sequence of final-SSB (token), SingleSeq = 0.

Step 2 Read final-SSB.

Step 3 If final-SSB (token) is empty then test is successful and return SingleSeq = 0.

Step 4 for each sequences of final-SSB (token) do

(a) If all ith items in every sequencing are same token t. And if total count of these tokens \geq Min up create a new item t with count and insert it into Single Seq.

(b) Otherwise, return false and Single Seq = 0. Step 5 Test is successful and returns SingleSeq.

Algorithm ConstructSubCSB:

Step 1 Initialize final-SSB(token+t) =0

Step 2 For each item in ei-queue of PSB(token+t), insert its suffix sequences into final-SSB(token+t).

Step 3 Return final-SSB (token+t).

After mining process, two text files are created named “Finalpattern.txt” for frequent access patterns and “URLOnly.txt” for Unique URLs. When a web user access a web page, if web page is present in “URLOnly.txt” then Recommendation module displays a list of recommended web page sequences. And, user will get knowledge about important web pages

The Pattern-tree; Creation parts constructs needed recommender model or Pattern-tree directly using mine sequence-access patterns. CSB-mine procedure and Pattern-tree creation procedure are conducted off-line. When customer visits website, customers HTTP demands in present surfing session are chronicled and present access sequence created. Matching customer's present current access-sequence from recommender design of Pattern-tree, the recommender rules creation part are generate recommender rules.

Procedure for Existing System

Input:

1: SAP – Some sequence-access Patterns

Output:

1. T – Pattern – SAP tree

Methods:

1. Create empty root node R for pattern
2. For every sequence S E SAP, represented as $S = e_1 e_2 \dots e_n$, do named e_1 , use maximum support existing between S and little node as support for e_1 and set present node point to e_1 create novel little node having current node point to novel little node.
3. Return said pattern tree.

Procedure in recommender rules creation

Input:

1. T – pattern-tree that considered threshold minSup
2. S – $a_1 a_2 \dots a_n$ – present access sequencing from customers
3. min-Length – minimum length for access sequencing
4. max-Length – minimum length for access sequencing that is lower than depth of pattern – tree

Output

RR- recommender Rule of ordered access events of S

Method:

1. Initialize RR - Θ
2. If $|S| > \text{Max-length}$, remove first $|S| - \text{max-length} + 1$
3. If $|S| < \text{Min-length}$, return RR else, set present-node point to root node R for T
4. For every item a_1 from head S to end do
 - a: if present-node possess little node named a_1 , then set present-node point to this little node
 - b: else eliminate first product from S and repeat from step 3
5. If present-node possess little nodes, insert involved little node to RR ordered by supports.

Class Diagram:

Unified modeling language class diagram is static structure which describes system structure revealing system classes, their features and connection between classes (Dhiman et al, 2012). The class diagram reveal foundation for object-oriented system. A class is represented with rectangle which is divided in three sections: class name, features and technique. The name for any class is, by convention, word with initial uppercase letter. It appears at first section of rectangle.

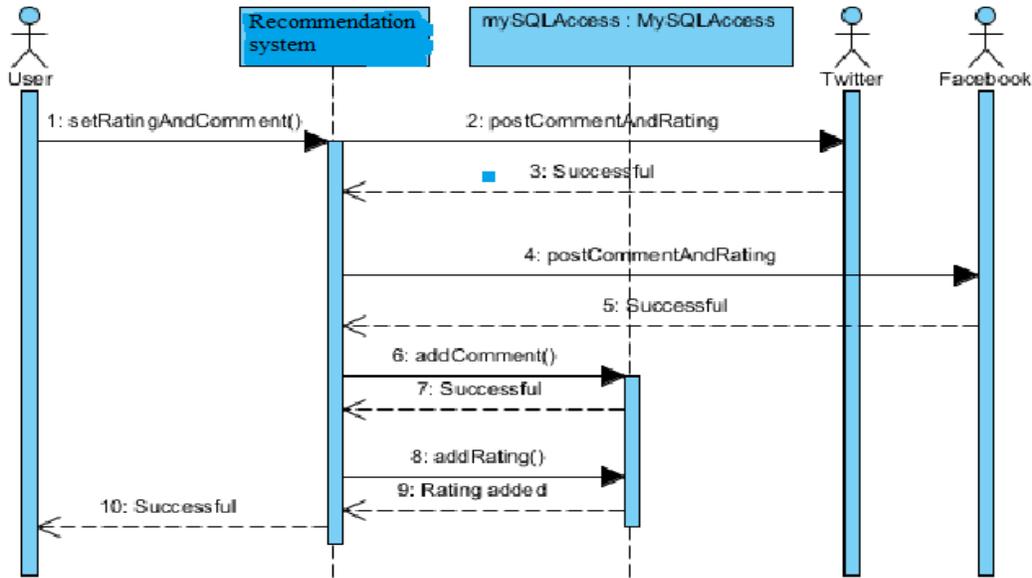


Figure 3.3: Sequences Diagram of the Proposed System.

3.4 METHODOLOGY

The overall software development phases which include analysis and design will implore occurring data and its performance, which depict in-depth knowledge of fragment allocation in distributed database System in system. The Object Oriented System Analysis and Design Methodology (OOSADM) was used in this approach.

4. RESULTS AND DISCUSSION

The comparative analysis of suffix and hybrid system was tested and deployed on the server after writing and debugging the program to ensure a convenient observation of the efficiency of the program is properly observed.

Experiment

Table 4.1 shows experiment that shows our program speed and effectiveness of suffix and hybrid searching for document like (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 4.1

Table 4.2: Experiment 2

Keywords	Existing system	Proposed system
Users profile (keywords)	0.94sec	0.84sec
Private information	0.64sec	0.7sec
Hybrid	0.82 sec	0.72 sec
Suffix	0.83sec	0.74 sec

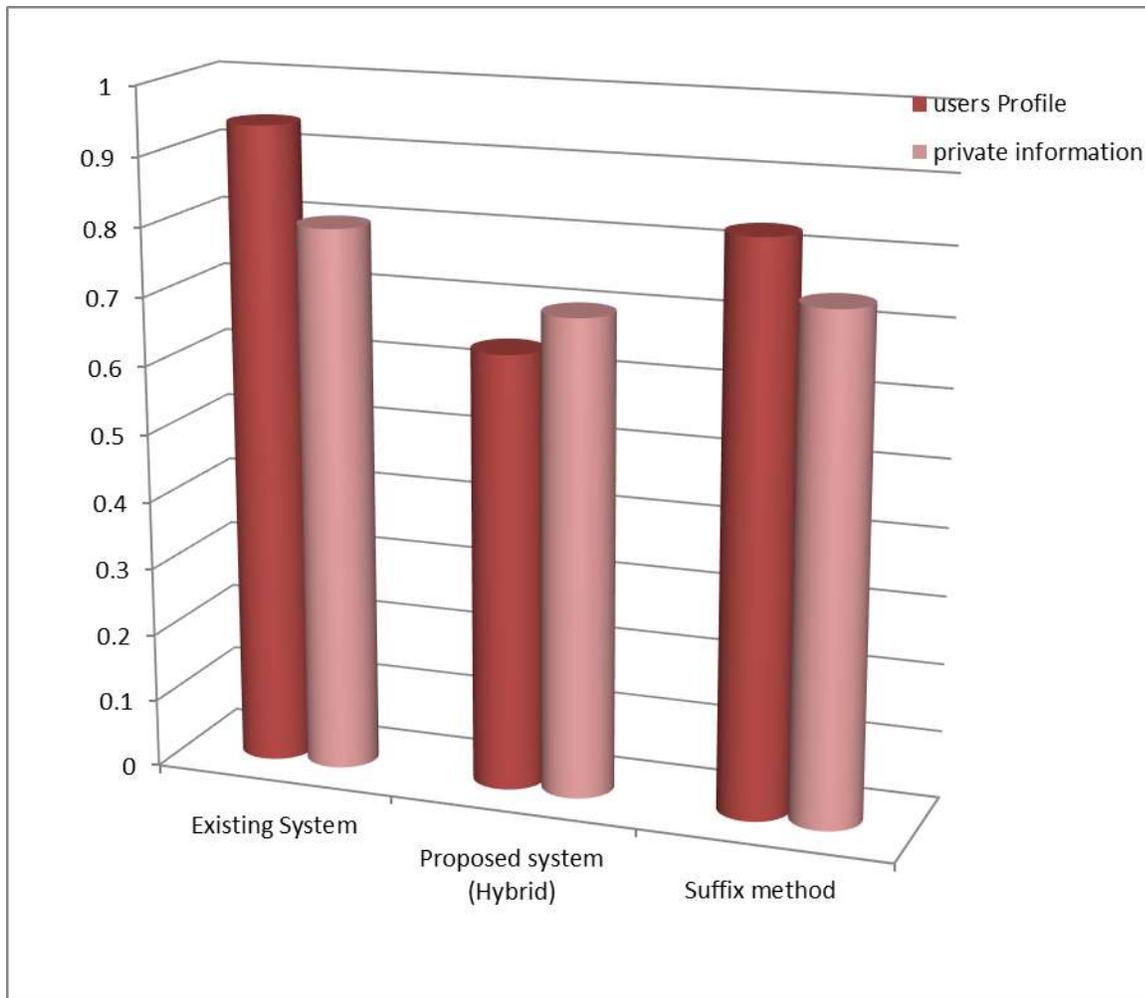


Figure 4.1 Hybrid and Suffix comparative graph.

5. CONCLUSION

Conclusively, this system uses Hybrid and suffix technique to provide effective search and rapid user token from the giving server to weblogers usage by comparatively analyzing different recommender techniques and token mining as text. The given system uses token mining to comparatively classify weblogers to suitable group and rate them in a categorize manner and collaboration filtering to group users with similar parameter and demographic profiles.

REFERENCES

- Baoyao Z,S.: Cheung H and Kuiyu C. (2004) ” An Intelligent Recommender System using Sequential Web Access Patterns” Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems Singapore, 1-3.
- Beel, J.; Langer, S.; Genzmehr, M.; Gipp, B.; Breitingner, C. (2013)."Research Paper Recommender System Evaluation: A Quantitative Literature Survey" (PDF).Proceedings of the Workshop on Reproducibility and Replication in Recommender Systems Evaluation (RepSys) at the ACM Recommender System Conference (RecSys).

- Bouneffouf, D. (2012), "Following the User's Interests in Mobile Context-Aware Recommender Systems: The Hybrid-e-greedy Algorithm", Proceedings of the 2012 26th International Conference on Advanced Information Networking and Applications Workshops (PDF), Lecture Notes in Computer Science, IEEE Computer Society, 657–662, ISBN 978-0-7695-4652-0
- Burkart, A., and Medlik, S. (1974). *Tourism – Past, Present and Future*. London: Heineman.
- Joeran B; Stefan L; Marcel G; and Andreas N. (2013). "Persistence in Recommender Systems: Giving the Same Recommendations to the Same Users Multiple Times" (PDF). In TrondAalberg; Milena Dobрева; Christos Papatheodorou; GiannisTsakonias; Charles Farrugia. *Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013)*. Lecture Notes of Computer Science (LNCS). **8092**. Springer. 390–394.
- Kumar, V., and Reddy, R. (2014). A Survey of Recommender systems (RSS) and its applications. *International Journal of Innovative Research in computer and Communication Engineering*, 2(8), 5254-5260.
- Lu, J., Wu, D., Mao, M., Wand W., and Zhang, G. (2015). Recommender Systems Application Developments: A Survey. *Decision Support System*, 74, 12-32.
- Shiva N, Mohammad H, and Ayoub B. (2011). A Hybrid Recommender System for Dynamic Web Users, *International Journal Multimedia and Image Processing (IJMIP)*, 1(1).
- Staab, S., Werthner, H., Ricci, F., Zipf, A., Gretzel, U., Fesenmaier, D., and Paris, C.(2002). *Intelligent Systems for Tourism*. *IEEE Intelligent Systems*, 17(6), 53-64.
- Su, W., Eichi, H., Zeng, W. and Chow, M. (2012). A Survey on the Electrification of Transportation in a Smart Grid Environment. *IEEE Transactions on Industrial Informatics*. doi:10.1109/TII.2011.2172454, 8(1). 1-10.
- Yong Ge; HuiXiong; Alexander Tuzhilin; Keli Xiao; Marco Gruteser; Michael J. Pazzani (2010). [An Energy-Efficient Mobile Recommender System](#) (PDF). *Proceedings of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. [New York City, New York: ACM](#). pp. 899–908. Retrieved 2011-11-17.
- Yoon Ho Cho, Jae Kyeong Kim, and Soung H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications* 23, 329–342.
- Zhou, B.Y. Hui, S.C. and Fong, A. C.M. (2004) "CS-mine: an efficient WAP tree Mining for Web Access Patterns mining for web access patterns", the 6th Asia Pacific Web Conference (APWEBW), Hangzhou, China, 523-532