Personalized News Recommender using Live RSS Feeds

Yash Dodia[1], Vaibhav Chauhan[2], Kiran Shewale[3], Aniket Darveshi[4], Komal Mahajan[5]

B.E. Student, Dept. of I.T., Atharva College of Engineering, Malad, Mumbai, India[1]

B.E. Student, Dept. of I.T., Atharva College of Engineering, Malad, Mumbai, India[2]

B.E. Student, Dept. of I.T., Atharva College of Engineering, Malad, Mumbai, India[3]

B.E. Student, Dept. of I.T., Atharva College of Engineering, Malad, Mumbai, India[4]

Assistant Professor, Dept. of I.T., Atharva College of Engineering, Malad, Mumbai, India[5]

ABSTRACT: Reading news online has become prominent as the web services provide access to news articles from various sources across the globe. Reading news has drastically evolved with the development of the World Wide Web (www), from the conventional way of reading a physical newspaper to access to numerous electronic news sources over internet. This project delineates a system that collects news from an eclectic mix of electronic news distributors. This system aggregates news from various HTML and RSS (Rich Site Summary) Web pages by utilizing source specific data extraction programs and tabulates them as per the predefined news groups to ameliorate personalized views through a Web based user interface. Our Personalized News Recommender System provides the users with access to the news articles gleaned from different Web sources. It takes the content from miscellaneous and heterogeneous online sources such as online news portals, bloggers, websites, etc. News items are categorized distinctly according to the user's tastes and the KEYWORD input given on the various categories of News articles.

KEYWORDS: News Recommendation, RSS Feeds, Online News, Keyword Searching, Personalization.

I. INTRODUCTION

A lot of people like to view and analyze news from various news sources. At instances people are looking only for the for the pinnacle news tales in their categories of interest. Hence are more likely to subscribe to RSS feeds from different news sources. Therefore, the users have to scan through all the top news stories in order to read stories of their likes and interests. Just like, a person interested in sports related top news category has to go through all the top news stories analyzing news from different channels. We identified this need of bringing together news from different sources and categorizing them and presenting them to the users as a single news feed.

Many users tend to subscribe to RSS feeds of their interest in order to get updated with the latest news. However, many a times this information is scattered across various news sources and spans more than one domain. Our system provides RSS feeds that presents all the news items from various news sources and groups them into categories with the main objectives such as:

☐ Providing the User with personalized news they like by analyzing the User's click behavior.

☐ Processing RSS feeds (representing news channels) and obtaining a single, well-categorized output feed.

☐ Analyzing and searching keywords entered by user to find a particular news article.

To make an application which responds quickly to user's action and preference and provide them better results. Its

major concern is to save time and find and categorize top news as per user's interest. Our project aims at processing multiple RSS feeds (four news channels) and obtaining a single, well-categorized output feed

II. RELATED WORK

This work [1] by Dr. M. Durairaj and K. Muthu Kumarv was carried out in June 2014.In this work, the first section discusses different News Recommendation System and its functionalities as well as the technologies involved in these systems. The second section discusses about extensively different topic analysis models and technologies involved to develop these models. The third section discusses the advantages and disadvantages of these systems as described by the respective authors. Finally, the paper is concluded with the suggestions and recommendations for building an effective news recommendation system based on the observations made from the extensive reviews and study. The observation made out from this literature study is that the news recommendation is challenging due to the rapid evolution of topics and preferences. The comparison and analysis are carried out on different techniques and methods used for mining the news recommendation models.

Disadvantages- The proposed system does not have good efficiency in terms of user keywords and search results. This work [2] was done by Florent Garcin, Kai Zhou, Boi Faltings and Vincent Schickel in 2012. This work considers 3 kinds of recommender systems: collaborative filtering at the level of news items, content-based recommendation where we recommend items with similar topics to what was read, and a hybrid where collaborative filtering is applied at the level of topics. Collaborative filtering recommends items to a user based on users with similar tastes, while content-based techniques create recommendations by analyzing the content of the items. Collaborative recommendation compares reading histories in order to extract reading behavior patterns. It recommends news items that other readers with similar reading histories have read. Readers are in different stages at a point of time, and news feeds are generated on basis of transition probability from one stage to another. In conclusion, it demonstrated that personalized recommendations using collaborative filtering can be useful even for individual newspaper sites with limited amounts of data about their users.

Disadvantages- The content-based and hybrid recommendations have surprisingly poor performance. This work [3] has been done by Xindong Wu, Fei Xie, Gong qing Wu and Wei Ding in the year 2011. In order to get real time updates of general Web news topics, interests and preferences of a user, a keyword knowledge base is maintained. The non-news content irrelevant to the news Web page is filtered out. Topics consists of keywords that are extracted using lexical chains which signifies semantic relations between words. Text summarization, collaborative filtering and recommendation of online news have received much attention in web and artificial intelligence, focusing on finding relevant and interesting news and also summarizing concise content. The Personalized News Filtering and Summarization system (PNFS) consists of two phases which are Personalized Web News Filtering and Web News Summarization. The purpose of keyword extraction is two-fold. First, it gives a concise form of the news to the user that saves the reading time. Second, the extracted keywords are also used to build a user interest model. In this work, it has presented the recommendation and summarization components of the personalized news filtering and summarization (PNFS) system. For the recommendation component, it has designed a content based news recommender that automatically obtains Word-Wide-Web (WWW) that is online news from the Google News portal and recommends news to users as per their preference.

Disadvantages - The Google news in Personalized News Filtering and Summarization system (PNFS) provides general news avoiding missing important news.

III. PROPOSED ALGORITHM

A. *User Module:*

In this module, the user runs the Application of Personalized News Recommender. In the application the user is asked to log into the Application or to register to the Application. After the User logs into the application, the user selects the News Category or searches news manually by entering keywords.

B. *RSS Feeds Module:*

In this phase the user selects the category of news of his/her choice.After that, RSS Feeds is used to fetch live news from various online electronic news portal.The RSS Feeds module sends back the results to Application Module where user can see the news recommended by the system.

C. *Web Server Module:*

This module is used to fetch each and every web page content which are stored in the database. User enters the keyword of his/her choice in the search box and web server is used. This module uses word weightage algorithm to fetch the weights of every document/web page in relation to the keyword entered by the user. Word Weightage algorithm uses TF-IDF technique to fetch the news article from the database. TF-IDF technique calculates the term frequency for every word of the individual web page stored in the database. By using this word weightage algorithm, it selects the document/web page of the highest weight and sends back the results to application module

D. *Algorithm Module:*

The algorithm used in this system is Word-weightage algorithm which is based on Tf-idf technique. *Term frequency-inverse document frequency*is abbreviated as Tf-idf, and the tf-idf weight is often used in retrieving data and text mining.

Normally, the Tf-idf weight formed by two terms: The first term computes the normalized Term Frequency (TF), which means how many number of times a word appears in a document, divided by the total number of words in that set of words or a document. The next term is Inverse Document Frequency which is abbreviated as IDF, which is computed as the logarithm of the no. of the files or docs in corpus, divided by instances i.e., number of files or documents where the precise term seems.

Hence this can be calculated by computing the following:

TF: Term Frequency, which measures how frequently a term occurs in a document. Since every document is different in length, it is possible that a term would appear much more times in long documents than shorter ones. Thus, the term frequency is often divided by the document length (aka. the total number of terms in the document) as a way of normalization:

TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document).

IDF: Inverse Document Frequency, which measures how important a term is. While computing TF, all terms are considered equally important. However it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus we need to weigh down the frequent terms while scale up the rare ones, by computing the following:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

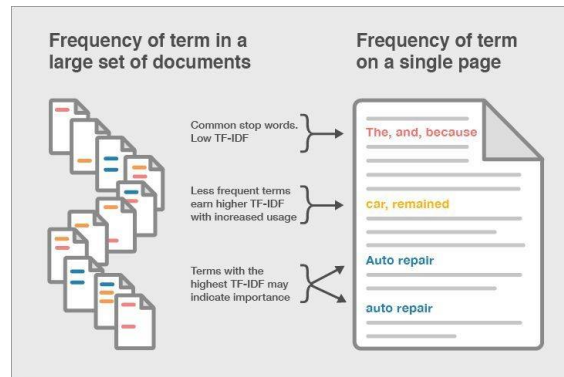IDF(t) = log e(Total number of documents / Number of documents with term t in it).

Fig.1 Algorithm Workflow

IV. RESULTS

We have developed a Web application for the proposed system. In our web application, the user will be first asked to register themselves [Fig.2]. During the registration phase, the user will be asked to choose the category of news of their choice [Fig.3]. After successful registration, the user will be asked to login into the application for viewing the news [Fig.4]. As the user logins into the application, all the news of user's preferred category will be shown on the home page of the application [Fig.5]. The user can view the news by clicking on view site button. The user can also share the article with other users by going on to the share tab of application. Over there, the user can see the articles that have been shared to them and also the articles which they have sent others [Fig.6]. In this application, the user also has an option to save the article offline and view the article later on [Fig.7]. If the user likes an article, then that article can also be marked as favorite and that article can be viewed later on by going into the favorite tab [Fig. 8]. The user can also view their profile by going onto the profile tab. Over there, the user can change and update their name, email id, phone number and password [Fig.9]. Thus, the Personalized News Recommender System works well as a web application and provides an accurate and quick online results of news and recommends it to the user.
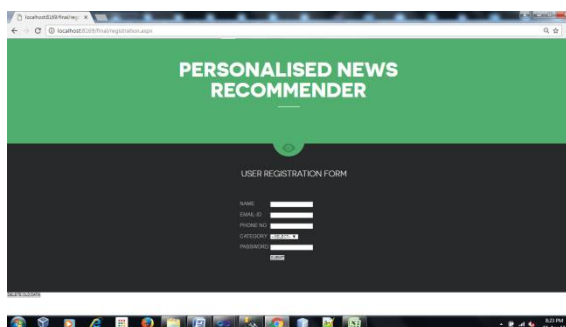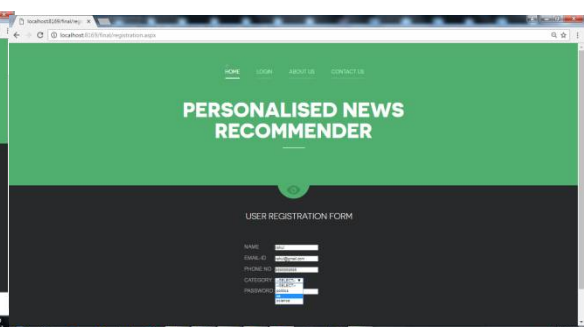

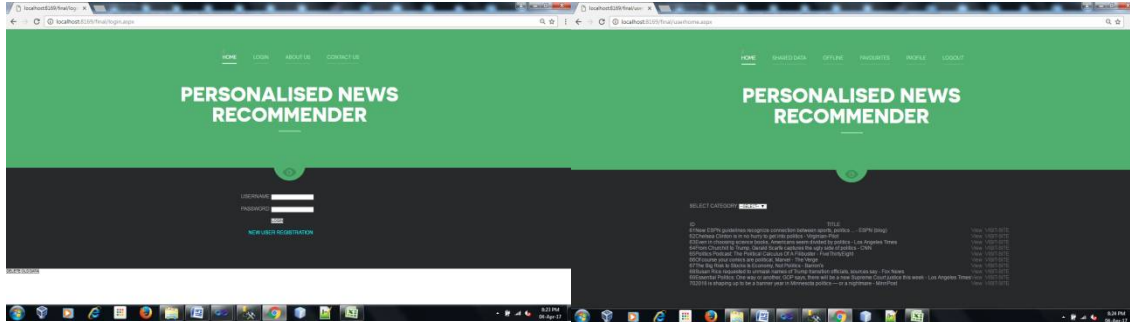Fig. 2 User Registration   Fig. 3 Category Selection

Fig.4 Login Page
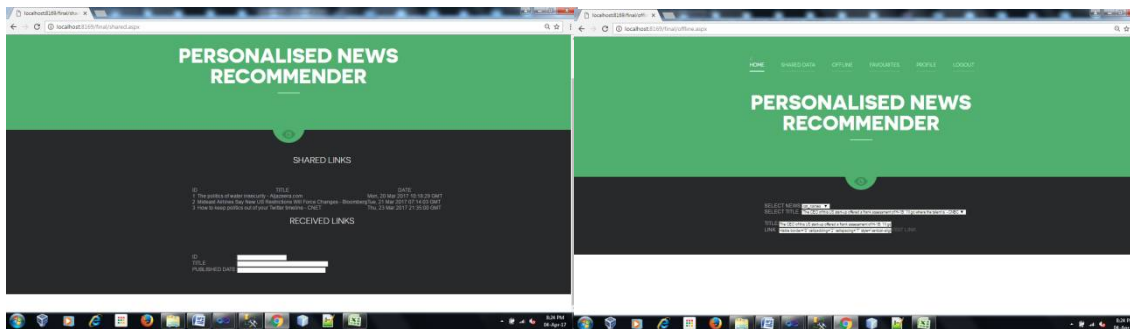

Fig.5 Home Page

Fig. 6 Article Sharing
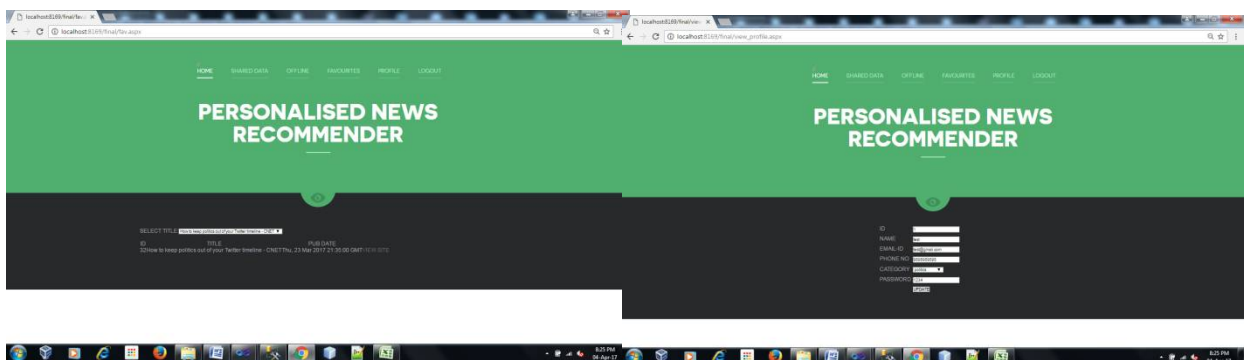

Fig. 7 Saving Article Offline

Fig. 8 Favourite Article

Fig. 9 User Profile

V. CONCLUSION AND FUTURE WORK

The personalized news recommender system works fine on RSS feeds and Keyword search using word weighted scheme. We ensure that the database updates are carried at regular interval of time. By giving Users what they like from the earliest starting point this application is expected to guarantee that individuals stay updated with what is going on around them.This framework gives a solitary RSS channel that introduces all the news objects from different diverse news sources and classifies them into different categories for the user to experience personalization. Previous recommender systems used to fetch the data from the already stored data sets and it was not necessarily live news.Thus we propose an application which recommends the right news at the right time to the user. Considering this system, use

of different languages and providing news according to the user preferred language can surely be a good work of future in this proposed system. Uploading the articles online can also be one of the future work in this system. Different attributes essential for the success of a recommendation system such as novelty, trust, coverage and privacy must be covered effectively. Cloud based approach for storing data can be used to improve the overall storage and processing of data.

References

1. Dr.M.Durairaj, K.Muthu Kumar"NewsRecommendationSystemsUsing WebMining:A Study" InternationalJournalofEngineeringTrendsand Technology(IJETT)Volume12 Number 6, June 2014.
2. Florent Garcin, Kai Zhou, Boi Faltings, Vincent Schickel "Personalized News RecommendationBasedonCollaborativeFiltering"IEEE/WIC/ACMInternationalConferences onWebIntelligence, 2012.
3. XindongWu, Fei Xie, GongqingWu, WeiDing"Personalized News Filtering and Summarizationon theWeb"23rdIEEEInternationalConference onTools. 2011.
4. Kyo-JoongOh,Won-JoLee,Chae-GyunLim,HO-JinChoi"Personalized NewsRecommendationClassifiedkeywordstocaptureUserPreference".12th International Conference on Advanced Communication Technology (ICACT),2014.
5. A. Balahur,R.Steinberger, M.A.Kabadjov,V.Zavarella,E. vander Goot, M.Halkia,B.Pouliquen,andJ.Belyaeva "SentimentAnalysisin the News" Proceedings of the7thInternational ConferenceonLanguageResources and Evaluation*(LREC)*, pp.2216-2220,May2010.
6. J.Liu,P.Dolan,andE.Pedersen,"Personalized news recommendation basedonclickbehavior"15thInt.Conf.onIUI,pp.31–40, 2010.
7. S.G.Esparza,M.P.O'Mahony,andB.Smyth"OntheRealtimeWeb as a Source ofRecommendation Knowledge", in *RecSys 2010*,September26-30 2010.
8. YuqiWang, WenqianShang"PersonalizedNewsRecommendationBased onConsumer's ClickBehaviour"12th InternationalConferenceonFuzzy Systems andKnowledge Discovery(FSKD),Vol.1,pp.634-638, 2015.
9. Jae-wookAhn, Peter Brusilovsky, Jonathan Grady, Daqing He, Sue YeonSyn" Open User Profiles for Adaptive News Systems: Help or Harm?" WWW2007 / Track: Browsers and User Interfaces, Vol.1, pp.11-20, 2007.
10. D. Billsus and M. Pazzani, "Adaptive news access," In: P. Brusilovsky, A. Kobsa, and W. Nejdl (eds.): The Adaptive Web: Methods and Strategies of Web Personalization, Springer, 2007.
11. A.S. Das, M. Datar, A. Garg, and S. Rajaram, "Google news personalization: scalable online collaborative filtering," in *Proceedings of the 16th International Conference on World Wide Web*, page 271-280. New York, USA, 2007.
12. K. Niemann , M. Wolpers , " A new collaborative filtering approach for increasing the aggregate diversity of recommender systems," Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 955-963, 2013.

BIOGRAPHY

Yash Dodiais a B.E. student in the I.T. Department, Atharva College of Engineering, Mumbai University, Mumbai, Maharashtra, India.

Vaibhav Chauhanis a B.E. student in the I.T. Department, Atharva College of Engineering, Mumbai University, Mumbai, Maharashtra, India.

Kiran Shewaleis a B.E. student in the I.T. Department, Atharva College of Engineering, Mumbai University, Mumbai, Maharashtra, India.

Aniket Darveshiis a B.E. student in the I.T. Department, Atharva College of Engineering, Mumbai University, Mumbai, Maharashtra, India.

Komal Mahajan is working as an Assistant Professor at Department of Information Technology, Atharva College of Engineering, Mumbai University, Mumbai, Maharashtra, India.