

Spring 2013

Recommendation System for News Reader

Shweta Athalye
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_projects

Part of the [Computer Sciences Commons](#)

Recommended Citation

Athalye, Shweta, "Recommendation System for News Reader" (2013). *Master's Projects*. 294.
https://scholarworks.sjsu.edu/etd_projects/294

This Master's Project is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Projects by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Recommendation System for News Reader

A Project

Presented to

The Faculty of the Department of Computer Science

San Jose State University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Shweta Athalye

May 2013

© 2013

Shweta Athalye

ALL RIGHTS RESERVED

The Designated Project Committee Approves the Project Titled

Recommendation System for News Reader

by

Shweta Athalye

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

SAN JOSE STATE UNIVERSITY

MAY 2013

Dr.Robert Chun

Department of Computer Science

Dr.Teng Moh

Department of Computer Science

Mr.Onkar Deshpande

Member of Technical Staff, VMware

ABSTRACT

**RECOMMENDATION SYSTEM FOR NEWS READER
APPLICATION**

By Shweta Athalye

Recommendation Systems help users to find information and make decisions where they lack the required knowledge to judge a particular product. Also, the information dataset available can be huge and recommendation systems help in filtering this data according to users' needs. Recommendation systems can be used in various different ways to facilitate its users with effective information sorting. For a person who loves reading, this paper presents the research and implementation of a Recommendation System for a NewsReader Application using Android Platform. The NewsReader Application proactively recommends news articles as per the reading habits of the user, recorded over a period of time and also recommends the currently trending articles. Recommendation systems and their implementations using various algorithms is the primary area of study for this project. This research paper compares and details popular recommendation algorithms viz. Content based recommendation systems, Collaborative recommendation systems etc. Moreover, it also presents a more efficient Hybrid approach that absorbs the best aspects from both the algorithms mentioned above, while trying to eliminate all the potential drawbacks observed.

ACKNOWLEDGEMENTS

I would like to thank Dr. Robert Chun, my Project Advisor for his guidance, support and faith in me, throughout the project. Thanks to my Committee members, Dr.Teng Moh, for his guidance in formulating the experiments conducted for the project and Mr.Onkar Deshpande for his constant feedback and suggestions. I thank my friends for their feedback and last but not the least; I would like to thank my husband, Mr. Saumitra Prabhudesai for his relentless support and faith in me.

TABLE OF CONTENTS

1.Introduction	8
2.Background Information	10
2.1 What are Recommendation Systems?	10
2.2 What are Mobile Recommendation Systems?	12
2.3 Content-Based Filtering Approach.....	13
2.4 Collaborative-based Filtering Approach.....	18
2.5 Hybrid Models	21
3.Project Design and Scope.....	23
4.Use Cases and Experiments.....	30
4.1 Use Cases:.....	30
4.2 Experiments:	33
5.Conclusion and Future Work.....	40
6. List of References.....	41
7.Appendix.....	43

TABLE OF FIGURES

FIGURE 1.RECOMMENDATION SYSTEMS	11
FIGURE 2.CONTENT-BASED FILTERING APPROACH.....	14
FIGURE 3.EXAMPLE OF CONTENT-BASED FILTERING APPROACH	15
FIGURE 4.CHALLENGES IN CONTENT-BASED FILTERING APPROACH	17
FIGURE 5.COLLABORATIVE-BASED FILTERING APPROACH	19
FIGURE 6.NEWSREADER APPLICATION ARCHITECTURE	23
FIGURE 7.NEWSREADER APPLICATION	29
FIGURE 8.RECOMMENDATIONS	32
FIGURE 9.CONTENT-BASED NEWS RECOMMENDATION RESULTS	34
FIGURE 10.COLLABORATIVE NEWS RECOMMENDATION RESULTS	35
FIGURE 11.COMPARISON RESULTS	35
FIGURE 12.USER DISTRIBUTION COMPARISON.....	36
FIGURE 13.HYBRID APPROACH RESULTS	36
FIGURE 14.HYBRID APPROACH V/S OTHER APPROACHES	37

1.Introduction

News- famously thought of being an acronym for the four directions-North, East, West, and South is an enriching source of conveying information on current events and trends presented to the readers through different source of media like print, television, internet and now with current trends in smartphones, in the form of news applications. Research on the news applications for Android suggests that there are around thirty-five applications [2] in the market which keep the users updated with the latest news across the globe. These applications provide news from most read newspapers like BBC News, Fox News, The Guardian, The Times etc. Various news aggregator apps are also in market, which bring multiple sources of news under one roof like Google Currents, Pulse News, Zite. These newsreader applications as we can see, focus on either particular newspapers or collate data from various news sources and present them to the user as per the current trends. While these applications are immensely popular with the masses, they can still be enhanced to provide personalized data to the users. Recommendation Systems filter large information sets from various fields of study, relevant to the interests of the users. They guide the users in judging various features and products made available to them by recommending the current trends. These systems form matrices based on various attributes and methodologies to provide suggestions to the users. These can be based on user preferences expressed either explicitly or implicitly or current trends (as in the application, Google Currents)[3].

The goal of this paper is to explore the various recommendation techniques used such as: *Content-based Filtering, Collaborative Filtering, and a Hybrid Model involving both*, to study the algorithms in detail and develop a newsreader application on Android Platform

using one of these techniques. The project aims at concluding which filtering algorithm is apt for various users.

The paper is organized as follows. Chapter 2 contains relevant background information. Chapter 3 presents the detailed description about the proposed technique for implementing the newsreader application on Android Platform. Chapter 4 discusses the Use Cases and Experiments conducted for the different Recommendation Algorithms. Chapter 5 summarizes the conclusions and gives an overview of possible future work. Chapter 6 provides the List of References. Chapter 7 is the Appendix, which includes the data set used for this project.

2. Background Information

2.1 What are Recommendation Systems?

Tremendous information about various products is available for the users and Recommendation Systems help in narrowing down these data sets and thus manage this information load by gauging the choice of a user. This is done by analyzing ratings, reviews and different such parameters which indicate the preference of a user for various items such as books, movies, news articles, restaurants etc. Primarily, two approaches have been devised, Content-Based Approach, which is based on the features of a particular product, and Collaborative-Based Approach, which considers sets of similar users, from similar backgrounds and interests. Examples of popular Recommendation Systems are various ecommerce sites such as Amazon, eBay and entertainment applications like Pandora Radio and Netflix. The components of a Recommendation Engine work are Data Collection, Recommendation Engine, and User Interface.

❖ Data Collection:

Explicit Data Collection includes the following:

- Collecting user ratings for a product in terms of most liked items to least liked items
- Presenting two items to an user and asking him/her to choose the better of them
- Analyzing the list of products liked by an user, this data can be collected from various advertisements of items enlisted on various social networking sites like Facebook, Google+ etc.

Implicit Data Collection includes the following:

- Listing the items that a user views/purchases
- Analyzing the number of times an item is viewed and the duration for which it is viewed

- Analyzing the user's social network and discovering similar likes and dislikes

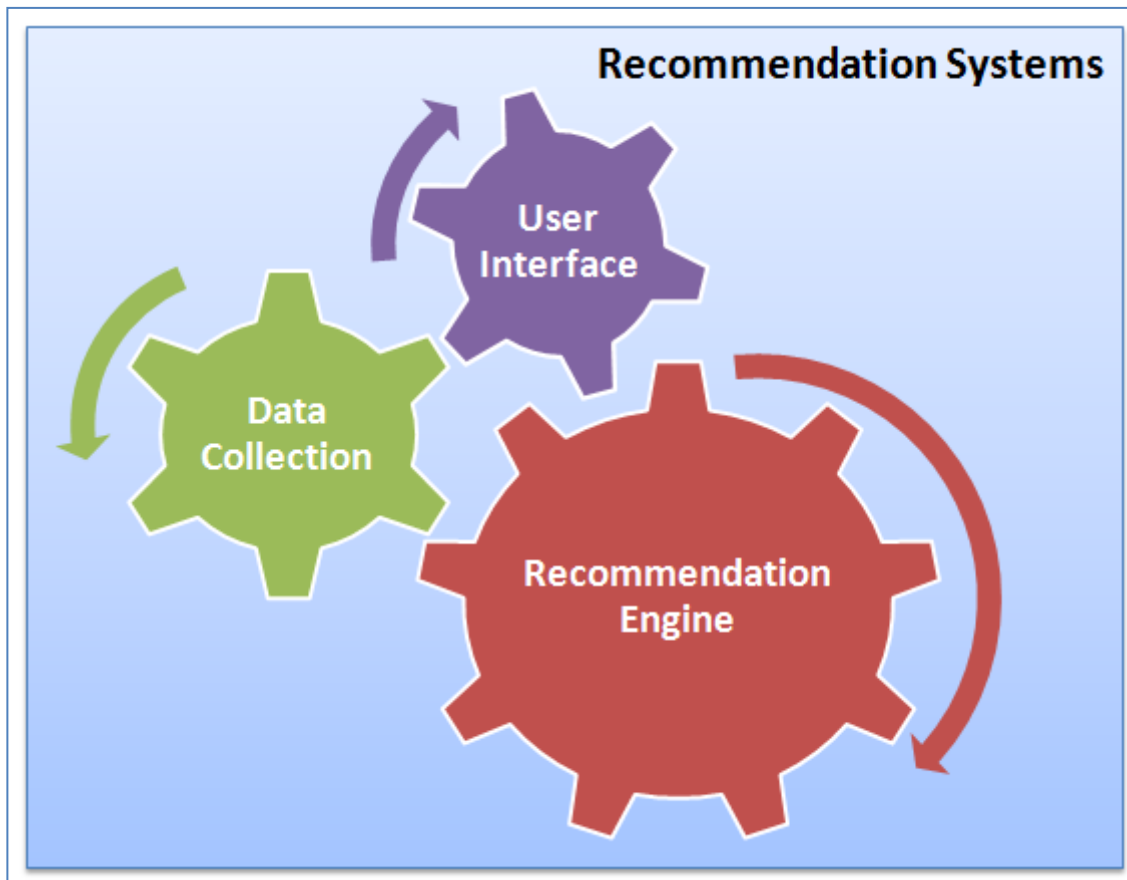


Figure 1.Recommendation Systems

❖ Recommendation Engine:

Recommendation systems are useful as they facilitate the users to gain access to data from a large pool of information, which they otherwise wouldn't have been able to find on their own.

Recommendation Engine is the brain of any recommendation system. Recommendation Systems are based on various approaches like Content-Based, Collaborative-Based and Hybrid Approach. Various algorithms like K-Nearest Neighbor, Pearson Correlation, and Rocchio Relevance Filtering etc. are used in the recommendation engine.

❖ User Interface:

User Interface is provided by Android Platform to give an enriching user experience, facilitating the users to view various items; news articles in this case.

2.2 What are Mobile Recommendation Systems?

With the advent of smartphone era, the competition has increased to present the data to users in such a way that it caters to their preferences as well as saves their time of specifically looking for information. Mobile Recommendation Systems provide assistance to users as they face decisions ‘on the go’. These systems face the challenge of making accurate recommendations using simple, yet appealing user interfaces [6]. A user browsing on a mobile application tends to spend less time than a user browsing on a desktop application. Hence, recommending items, which will hold the user’s interest in a short span of time, is one of the most important goals, which a mobile recommendation system should achieve. Most of the mobile recommendation systems heavily rely on locations of the user to recommend items to him. A person looking for restaurants in a particular location in San Jose, CA will find a recommendation system beneficial only if it gives suggestions relevant to his precise location. This holds true for various recommendation systems for shopping, hotel reservations etc. which heavily depend on the context in which the user is present. Social Networking applications like Facebook and Foursquare are used by users to check in their locations. They can also be used for recommendations and as review applications since a user checking in at a particular location can share his feedback about the place with a number of other users. The popularity of different locations, restaurants, and shops can be based on these feedbacks and used for recommendations [13]. A few other Recommendation Systems rely on popularity of the items among a large group of users. Recommendation systems for movies, books, and news articles consider the popularity of these items as preferred by a large group of users, irrespective of the location of the user. Users prefer watching the most

popular movie, reading the most popular book by an author and so on. The popularity of the news articles can be judged on the basis of the coverage of current breaking news across the globe or at the location where the reader is currently present. At the same time, a reader can prefer reading articles only from a set of newspapers he likes to read daily, irrespective of his current location or popularity of the articles. For example, a reader currently residing in US can still prefer reading news from his favorite newspaper based in Mumbai. Given such a case, it is important that the recommendation system keeps in mind the preferences of the users to personalize information for the user using either implicit or explicit data collection techniques. User's interest in various news categories can differ with factors like Recency, Novelty and Criticality of News Articles. For e.g.: During the outbreak of the hurricane "Sandy", number of users reading articles about aftereffects of Sandy as well as factors causing such hurricanes must have increased multifold. Many users must have read articles on Cyclones and related environmental issues.

2.3 Content-Based Filtering Approach

Content-based Filtering, as the name suggests, relies mainly on the content of the items in a collection. It works by mapping similarities between different items present in an information dataset. It does not consider the preferences or activities of other users. When a user shows that he likes a particular item by either purchasing it, or by giving positive reviews or ratings to it, the Content-Based Recommendation System maps it with similar items. Items with high degree of similarity are then presented to the user as recommendation. Various features of an item are considered in order to establish the similarity between these items. Thus the grouping of items is done based on the genre and subject matter [7]. For instance, a collection of various articles can be compared based on various categories like Fiction, Adventure, Comics, Science, Reference Books, Journals, Biographies and Autobiographies, Author Names, Publication House, Prize-Winning Books, Reviews, Ratings. These factors together

can be used to judge the similarity between items. The figure below illustrates the Content-Based Filtering Approach.

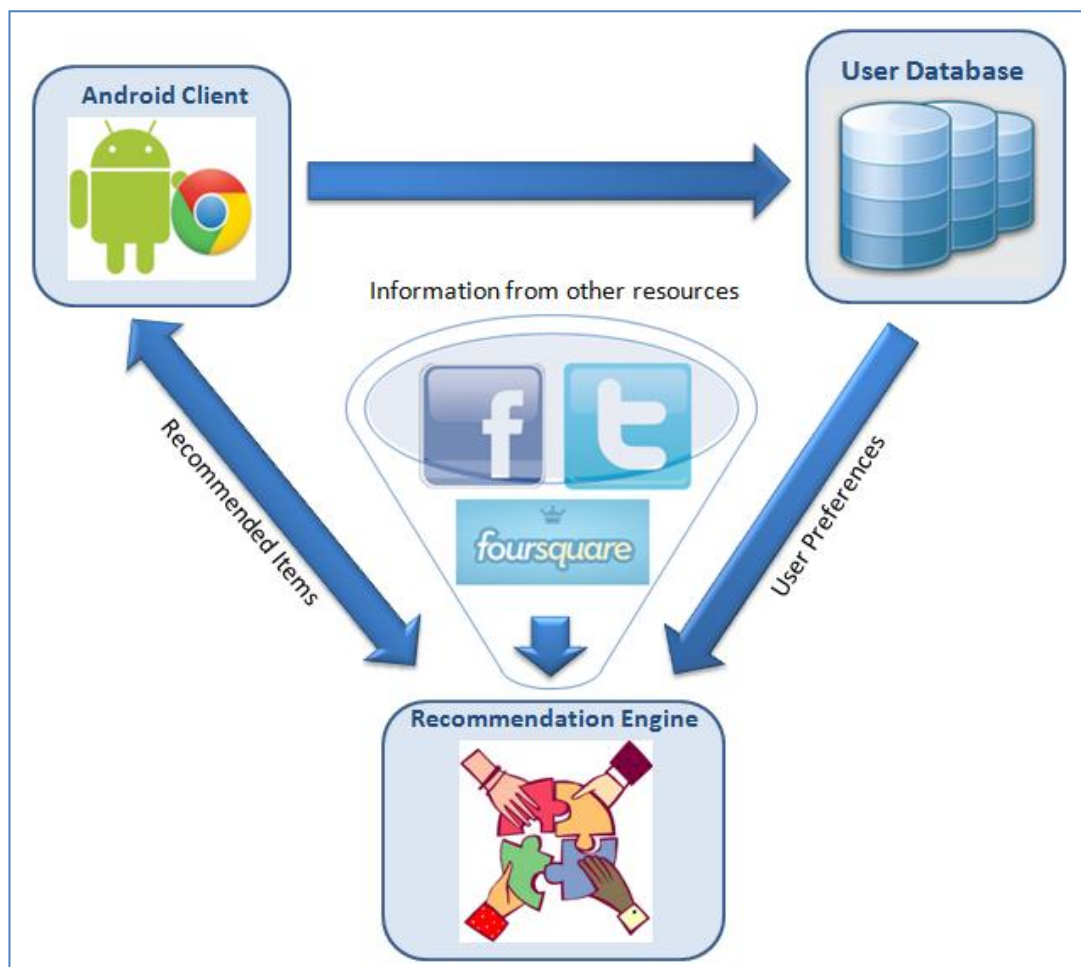


Figure 2.Content-Based Filtering Approach

The figure below depicts how Content-Based Filtering works in a real time scenario.

Amazon.com is a popular e-commerce site and is well known for the recommendations it makes to the users. The screenshot below shows an example, where a user who is a Computer Science Student, purchased a fiction novel, “The Immortals of Meluha”. The next time the user logs in; he sees a list of recommended books as below.

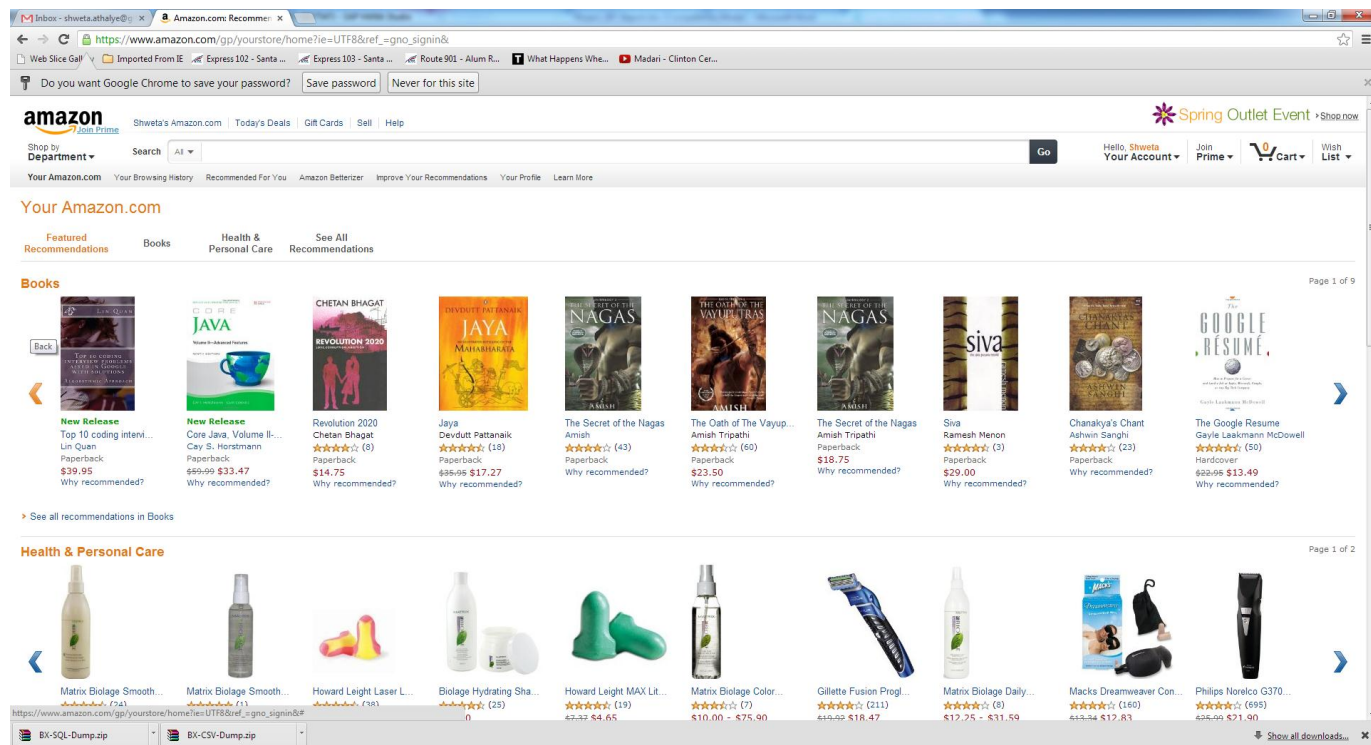


Figure 3. Example of Content-Based Filtering Approach

This book, written by an author named Amish, is a part of a trilogy with two more books, which are a part of the Recommendation List. The list also has books based on Mythology similar to the one, which he had bought. The user is a Computer Science student and has previously purchased books on Java, Interview Questions; hence the suggestions for books like “Core Java-By Dr. Cay Horstmann” are also relevant to the user. Thus, a content-based system encompasses a larger spectrum of attributes while making recommendations to a user, rather than just returning his searched results. Content-Based Algorithms use Text Analysis method, which considers lexical similarity between keywords. Ontology-based recommendation systems consider Semantic-level similarity [14]. Relevance feedback is an important feature of Content-Based Filtering approach as it helps in eliciting response from the user and understanding whether a user likes or dislikes a content [9].

Challenges faced by Content-Based Filtering are:

- (1) Overspecialization: Overspecialization is usually caused since Content-Based Filtering Approach relies on the features of the items. Comparing items on the basis of limited features can result in repetitive results. The recommendations can be too similar to the items already liked by a user in the past [16].
- (2) Limited Content Analysis: Content-Based system has its roots in the study of various attributes that define an item. In cases where materials are easily indexed by computers (digital articles, websites, etc.) or where rules and standards explicitly define the attributes associated with each item (library holdings) this is usually not a problem. However, multimedia, videos, music, and pictures can be much more difficult to accurately describe, and often these descriptions are subjective to the person doing the cataloging. Without accurate methods of comparing these types of materials, content-based filtering is ineffective. A user searches for an article: “User attitudes towards news content personalization”. Apart from displaying the requested article, the website also gives a few suggestions which are of no relevance to the user.

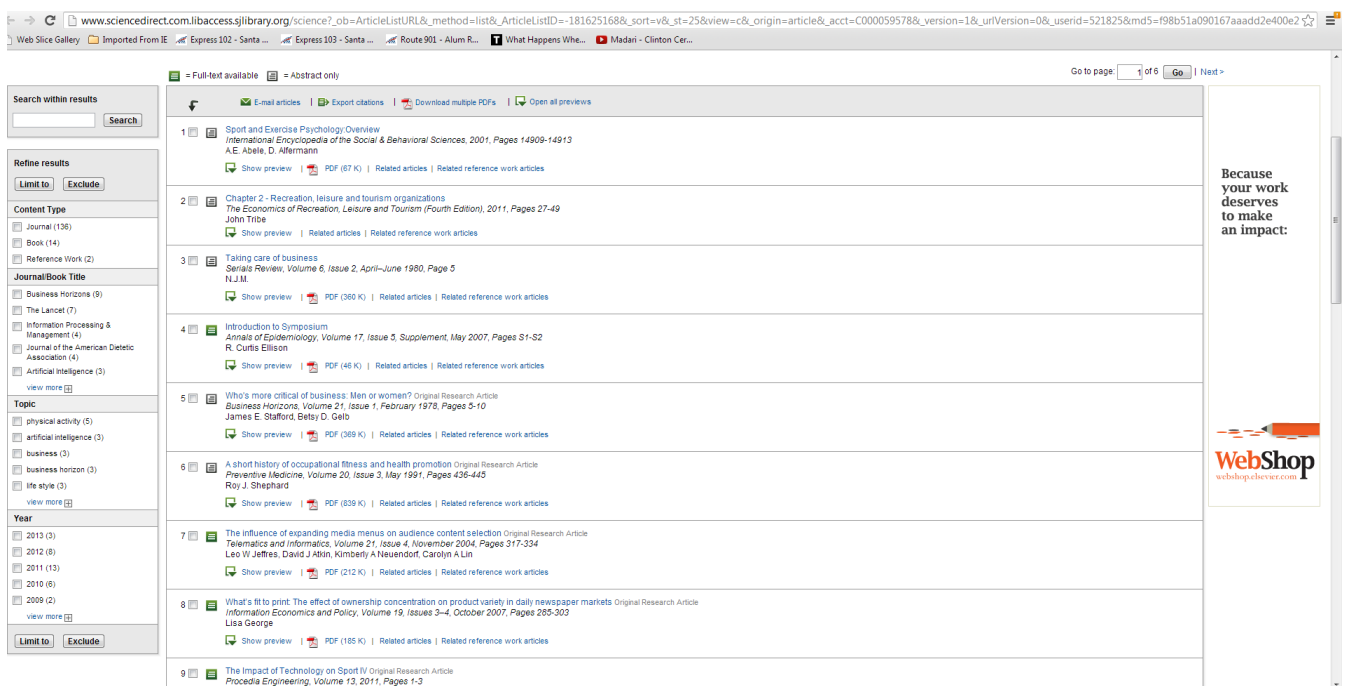
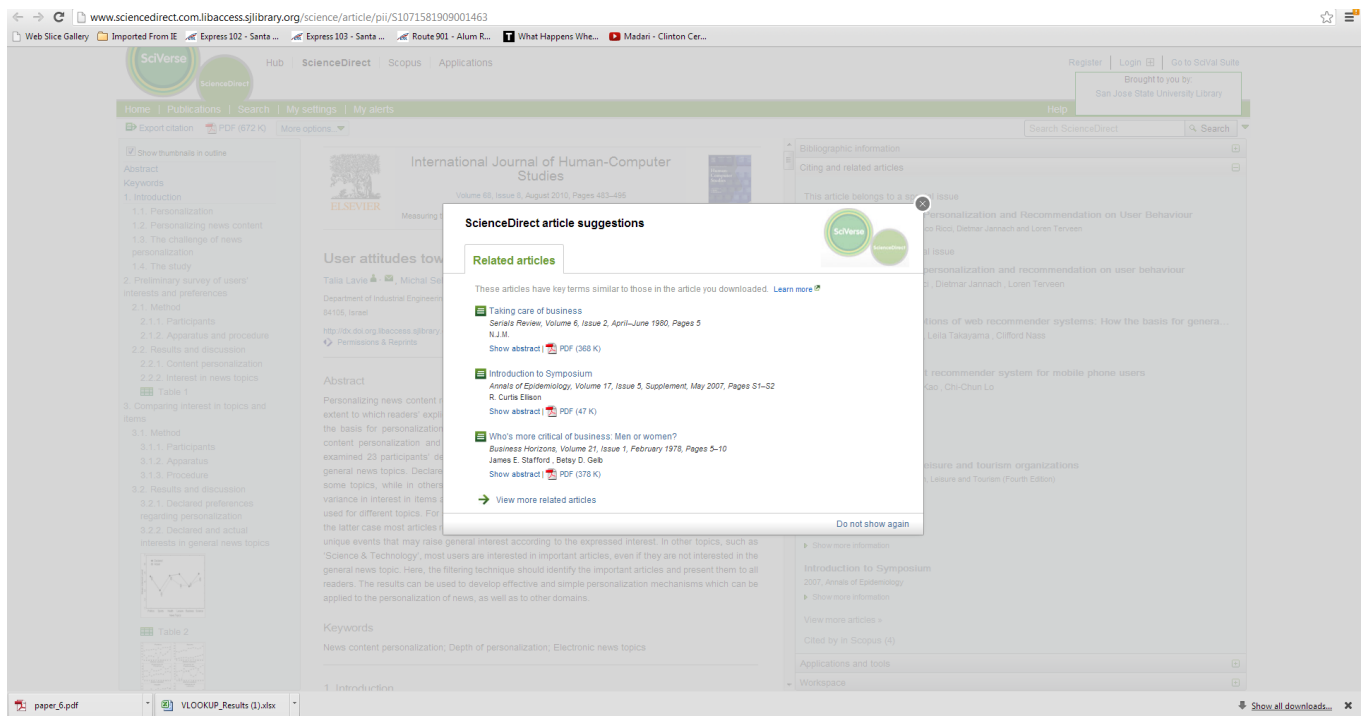


Figure 4. Challenges in Content-Based Filtering Approach

(3) New Users: User Profile Information collected requires considerable amount of data and features to define an item. Adequate information is thus required to be able to recommend effectively [17].

2.4 Collaborative-based Filtering Approach

Collaborative-based filtering approach, which is based on historical information of users, has been the most successful recommendation system approach and has been widely used in various applications. The first system to generate automated recommendations was the GroupLens system, which provided users with personalized recommendations on Usenet postings. [15]. Collaborative filtering is normally of two types:

- (1) **User-Based:** As the name suggests, this approach is based on users rather than the characteristics of items that a user prefers. It assumes that similar users have similar preferences. Similarity between users is based on various attributes like age, sex, and occupation etc.; which normally tend to define an individual's preferences and behavior. "K" most similar users (neighbors) to the given user are selected based on the factors as stated above. According to the definition of Collaborative Filtering, the historic information of the preferences of similar neighbors to a given user is studied and the top-N items in the list, which have not yet been purchased or used by the given user, are returned as recommendations.

The problems faced by user-based collaborative filtering systems are Scalability, Sparsity, and First-Rater Problem or Cold Start [15].

- (2) **Item Based:** Item-based approach considers the similarities between items used by users in the same group. Thus, similar items used by K-neighbors of a given user are used to generate recommendations using cosine similarity, excluding the items already used by the user [15].

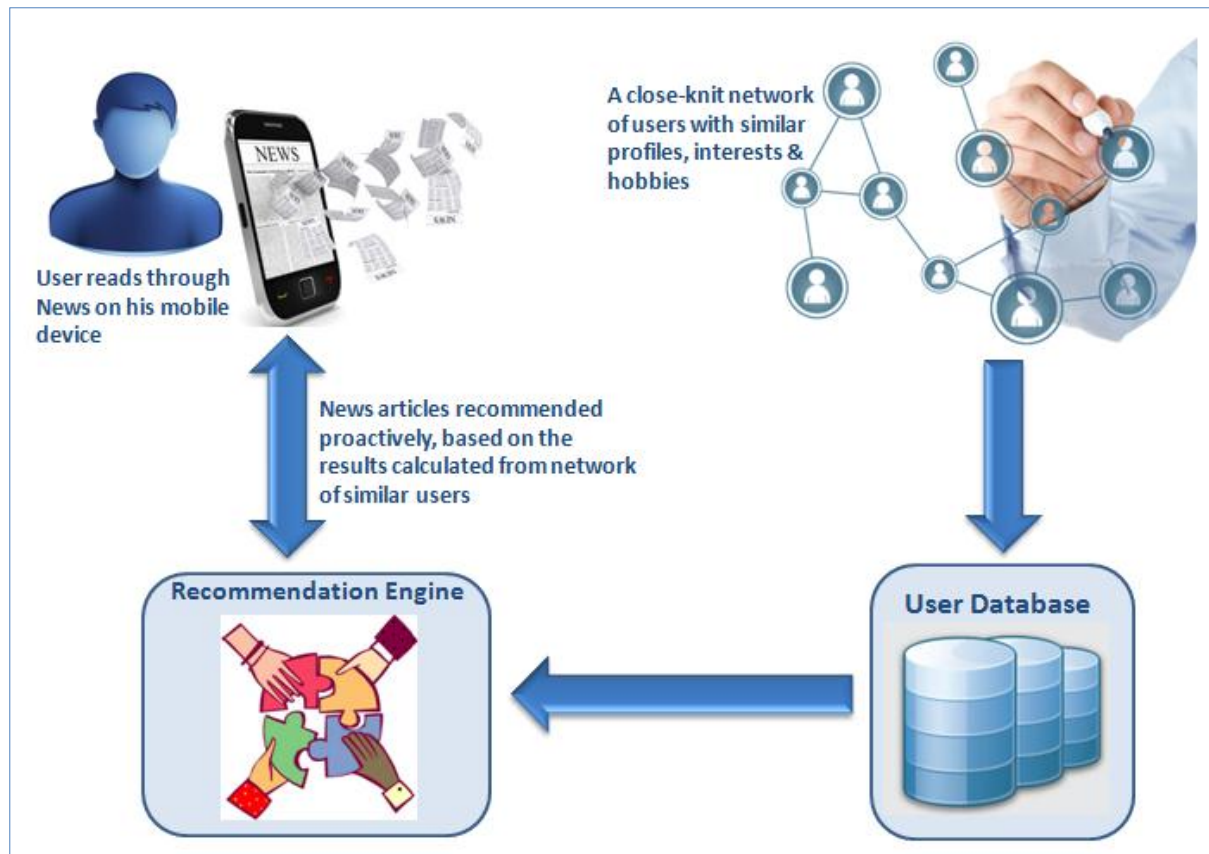


Figure 5. Collaborative-Based Filtering Approach

Challenges faced by Collaborative filtering are:

- (1) **Sparsity causing Cold Start (new user or new item problem):** Collaborative filtering relies on very large information sets. The user-item matrix used for such recommender systems can be sparse posing a challenge to the overall performance of a recommendation system based on Collaborative filtering. Cold Start indicates that a recommendation system will take considerable amount of time to form a preference list of a new user after he starts using a system. Thus, despite of having a group of similar users, the ratings obtained can still be sparse.
- (2) **Scalability:** Collaborative filtering essentially depends on the number of users of a system. As the number of users increases, the complexity of analyzing similar users and examining the history of the items used by these users increases multifold.

Scalability is also an issue where online systems need to react immediately to requirements of users [5].

- (3) Synonyms and Similar Contexts: Recommender Systems can fail to understand the similarity between synonyms like: Movies and Films. This can reduce the performance of the systems. Techniques like Topic Modeling can be used to group words with the same meaning to make recommendations effectively [5].
- (4) Gray Sheep: This refers to the users whose interests are dynamic and hence benefit less from a collaborative filtering based recommendation system, which makes suggestions based on the preferences of similar users [5].
- (5) Shilling attacks: Recommendation Systems, which explicitly collect data from users in the form of ratings and surveys, can be impacted negatively due to biased ratings given by users. Brands, which are cheap in quality and pricing and thus, popular with the masses, can get more ratings. This can have an impact on the correctness of the recommendations. Also, users can give positive ratings to the products used by them and negative ratings to those, which are not even used by them! Analysis of these preferences can thus lead to misleading results [5].
- (6) Diversity or Coverage: Collaborative Filtering approach brings about diversity and increased coverage while recommending items as it analyzes historical behavior of a large group of users. However, items, which are new or have limited ratings, are not covered in the process, which can lead to less diversity. Also, only if users have rated the exact items, they are considered as similar users. Therefore, if one user liked the Yahoo Finance page and another liked the BBC Finance page, the two would not necessarily end up being nearest neighbors [9].

2.5 Hybrid Models

Problems to both the systems are eliciting user feedback. Rating documents is a burdensome task for users, so fewer ratings are desirable. Ironically, for systems using a pure content-based approach, a user's own ratings are the only factor influencing future performance [9], and reduced ratings would mean inaccurate results and reduced performance owing to insufficient data. Item-Based Collaborative Approach for Recommendations has been devised to improvise the User-Based Collaborative Approach and has been successful in overcoming its flaws. In one model, content-boosted collaborative filtering is used to deal with the problems of new items and sparsity [16]. Essentially, this method works well in systems, like libraries, movie sets where the number of items is relatively high and there is a high probability that not all the items are rated by the users. A Hybrid System is the one, which takes the average of the ratings generated by pure content-based and collaborative-based approaches [7]. Both the Filtering Approaches, Content-Based and Collaborative-Based can be successfully used in various recommendation systems, if they are tailored appropriately to cater to the features of the application in use. It can, however be proved empirically, that the hybrid methods can provide more accurate recommendations than pure approaches as they can overcome the problems faced by these approaches as discussed above. Netflix is such an example of a system, which uses hybrid approach for recommending movies. It uses the history of the user's search results to compare with similar sets of users, which is Collaborative-filtering and also looks for movies which are similar in content to the historical search of an individual user, which is Content-Based Approach.

This project uses the Content-Based Collaborative-Hybrid Approach. Popular news articles read by various users in a dataset were initially displayed to the user. Ratings given in the form of clicks by the user were considered as user's preferences. No other personal information was recorded. A concept of Predictive Ratings was introduced for the content,

which is not rated by a particular user. A pseudo rating was assigned based on the ratings of the user for other items. Thus, similar content was rated similarly. In the process, the users with maximum ratings (actual plus system generated) have precedence over other users. These ratings were then used normally when comparing user profiles for collaborative filtering.

3.Project Design and Scope

The Project aims at including the features of both Content-Based and Collaborative-Based Filtering Approaches, considering the advantages of the Hybrid Approach and tries to improvise over the shortcomings of both the approaches to provide proactive recommendations to a newsreader. The main components of the system are:

- (1) Parser
- (2) Recommendation Engine
- (3) Database
- (4) Android Client

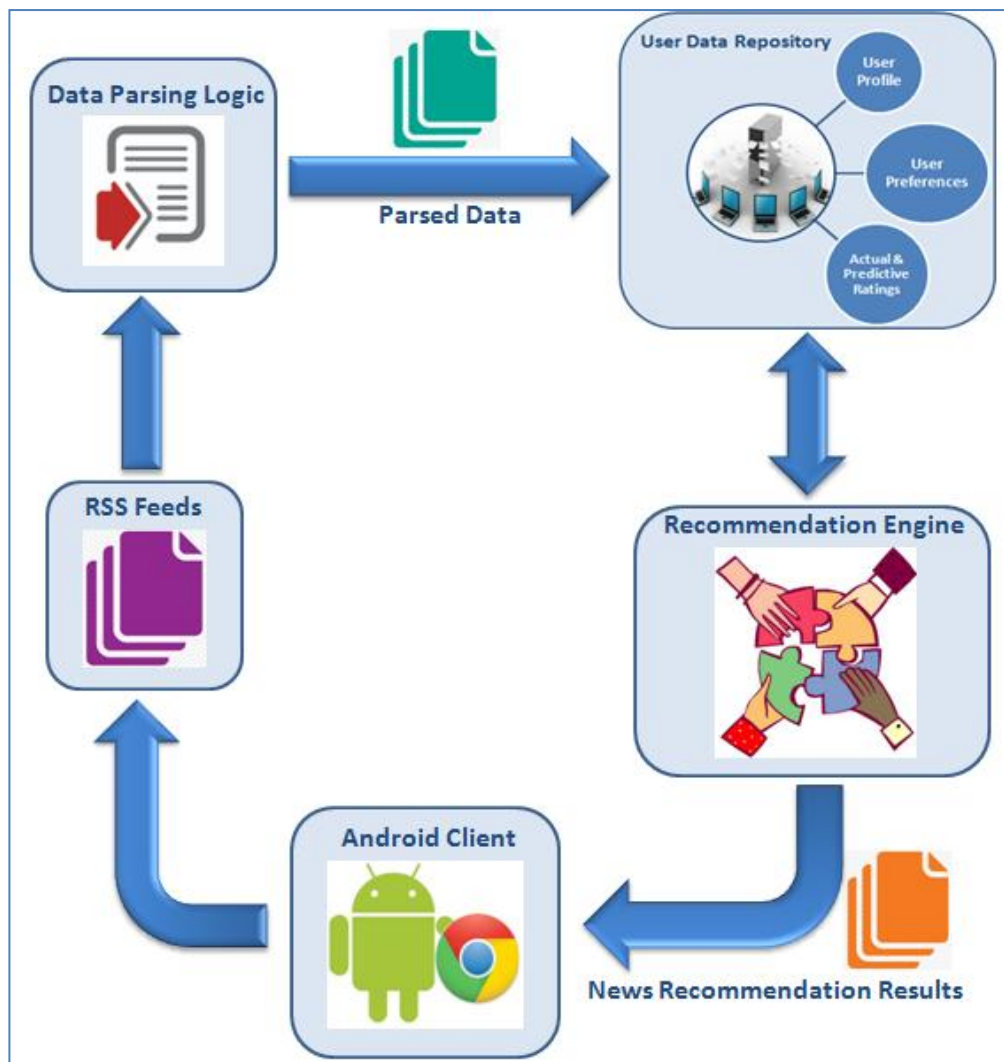


Figure 6. NewsReader Application Architecture

- 1) **Parser:** Data is parsed from the web for different news articles from the RSS feeds provided. Apache HttpClient library is used, the pseudo code of which is as below:

```
HttpClient oClient= new DefaultHttpClient ();  
HttpPost oPost= new HttpPost ("http://10.185.212.252/~shwetap/mastersproject/get_all_articles.php
```

The data will be categorized as following:

- If the user reads an article from Yahoo-Finance, the parser should break down the information as News Article belonging to Yahoo and Category-Finance.
- Android provides API for recording the events. So, for every click on the link, the application can store the name of the newspaper and the section or the category read.
- Weights can be assigned in order to calculate the ratings

As shown in Figure 6.the Parser stores the data in the form of a document, which is retrieved by the recommendation engine.

- 2) **Recommendation Engine:** The recommendation engine that provides the recommendations based on the historical information of the user over a period of time has been developed in Java using a hybrid approach involving both Collaborative and Content Based Recommendation Algorithms. It analyses the data from the parsed RSS feeds and performs the following steps:

Working of the Recommendation Engine:

- (i) Parse the RSS feeds of popular newspapers to get the information about urls and the descriptions of news articles based on Collaborative Filtering Algorithm
- (ii) Record the number of times a reader clicks on the links displayed on the newsreader application
- (iii) Analyze the score of the number of clicks for a month starting from the first time the reader used the application.
- (iv) Calculate the recommendation for news articles based on the weights assigned to the news articles as per the user's scores for a month and so on
- (v) The RSS feeds consist of the "title"-which tells us about the newspaper and the category being read, "url" of the news feed, the "pub date"-Date on which the article has been read. The urls which the users click are processed as follows:

Newspaper	Category	Score (No. of Clicks)	Date
Yahoo	Business	2	3/4/2013
Yahoo	Astrology	2	3/4/2013
Yahoo	Economy	1	3/5/2013
CNN	Business	2	3/5/2013
TOI	Entertainment	4	3/5/2013
Yahoo	Finance	2	3/5/2013
CNN	TV & Video	1	3/6/2013

- (vi) As mentioned earlier, Synonyms and Similar Contexts is one of the challenges faced by Collaborative Filtering. I try to overcome this challenge by considering similarity in the Categories read by the user. For instance, Economy is a subset of the Business section. Movies, Entertainment, TV & Videos fall in the Entertainment Section. The Recommendation System here

uses the Content Based Filtering Approach to study the forms or content of the items; newspapers and the categories in this case

- (vii) Thus, the processed data for a particular date range now looks like this:

Newspaper	Category	Score (No. of Clicks)
Yahoo	Business	5
Yahoo	Astrology	2
CNN	Business	2
TOI	Entertainment	4
Yahoo	Entertainment	3
CNN	Entertainment	3

- (viii) The Recommendation System uses the weighted combination of the scores of the attributes; Newspaper name and Category and displays the news articles to the users in the order of their occurrence
- (ix) The Recommendation System thus implements a hybrid model consisting of both, Collaborative and Content based approaches to cater to the needs of a newsreader. The data collection is implicit since the user need not disclose his personal information, preferences to the system
- (x) Predictive Ratings are given based on the actual ratings to ensure uniform distribution of the recommendation results.

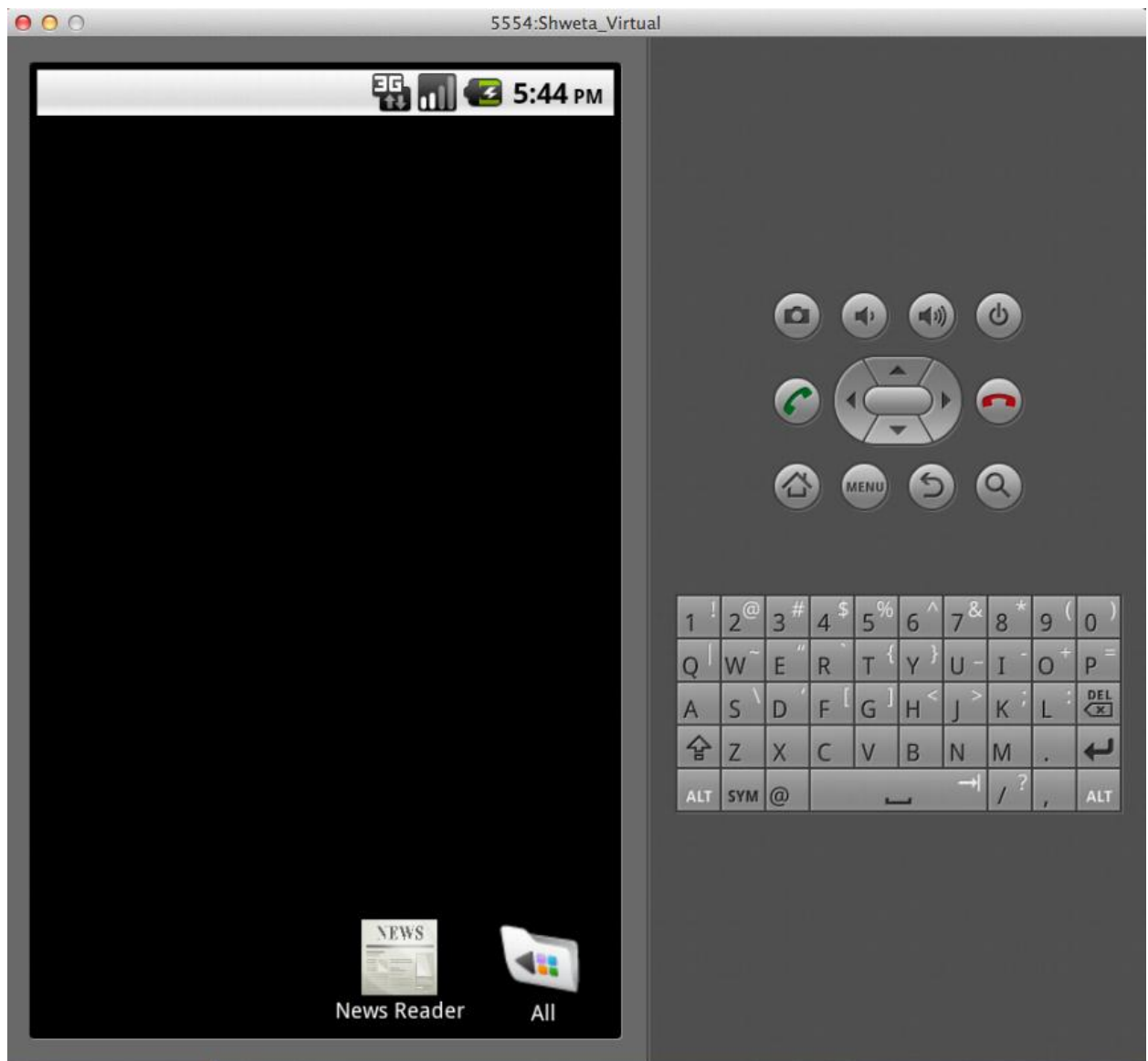
Calculate the recommendation from a weighted combination of the required preference, his past relevant information and his current location.

3) Database: MySQL has been used, as the data to be stored is user centric. The database has two tables, one for storing the classification based on the urls such as the Newspaper name and Category along with the username and timestamp. It also store the weights assigned to

the newspaper name and the category combination depending on the number of times the user reads these news articles. This is done by recording the number of times a particular url has been clicked by a user.

4) Android Client: Android APIs are used to develop the client application and the user interface. Currently android is the fastest growing mobile operating system and thus the application will appeal a large group of users. The main goal of the recommendation application for news articles is to proactively make recommendations to a user and personalize them based on his activities, recorded implicitly.

To cater to these needs, the newsreader application has a simple interface, displaying the news articles of various categories such as Finance, Sports, and Movies etc. from popular newspapers.



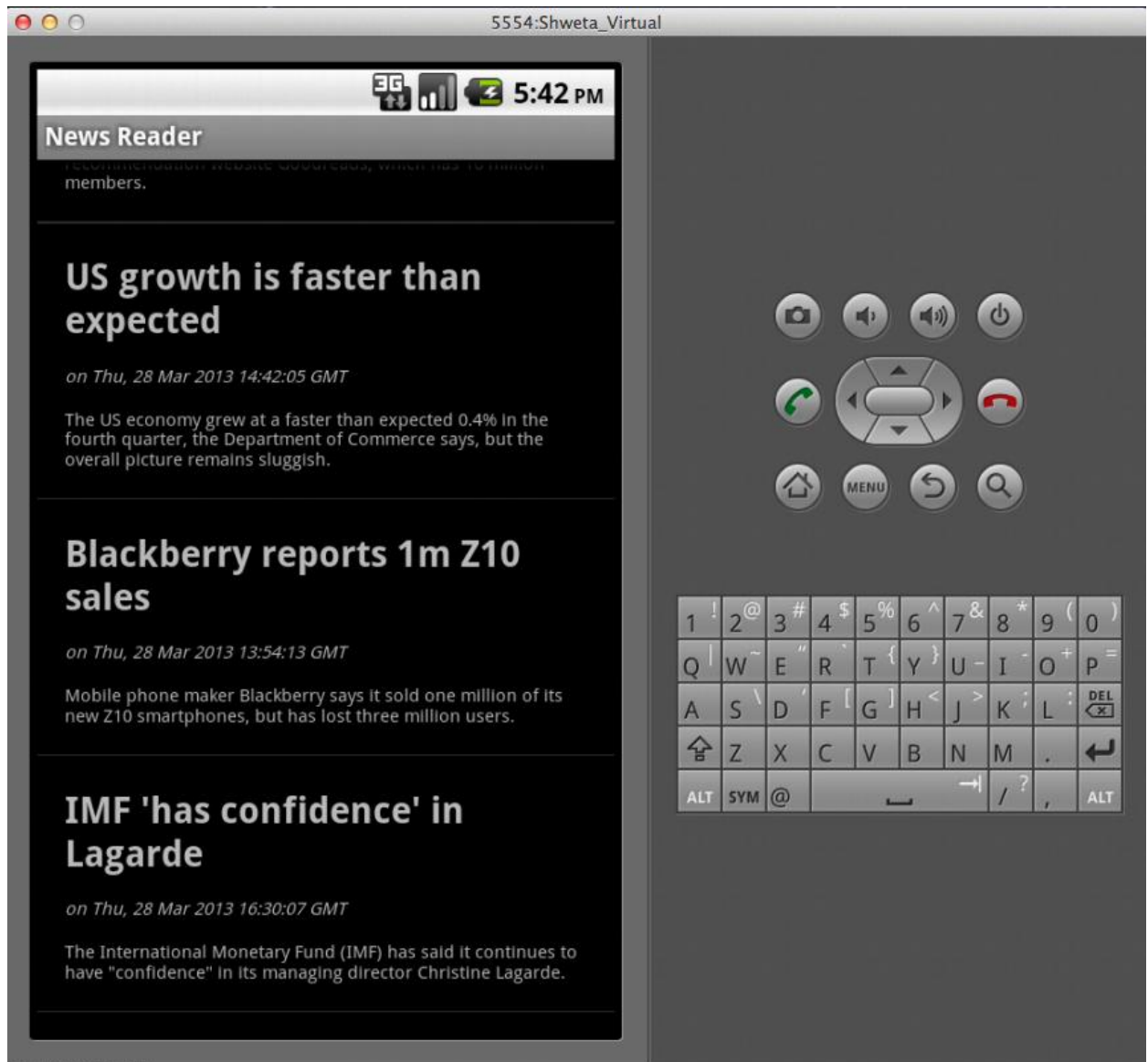


Figure 7. NewsReader Application

The user activity over a period of time, example, a week is analyzed and the news articles are displayed in the order of their maximum occurrence. Thus, user gets to read news of various categories from different newspapers, yet the application does not act only as a news aggregator since it recommends the articles proactively to the user.



4. Use Cases and Experiments


4.1 Use Cases:


The application is completely user-centric; hence the application focuses on providing news articles to the users as per their individual preferences (Content-Based Filtering) as well as depending on the similarities of reading patterns in a group of users (Collaborative-Based Filtering). The brain of any Recommendation application is the Recommendation Engine and the algorithm, which is used to implement the same. As discussed in this paper earlier, weighted approach has been used for recommending the news articles to users. The user interface has news articles from various categories like Sports, Finance, Politics, Movies, and Music etc. Various criteria [11] such as user's Gender, Age Range, Occupation Status, Socio Economic Status, Location can be pivotal in deducing conclusion about the user's reading pattern. However, these criteria can be misleading too. For instance, a person who is an Economist reads books on Economics, keeps himself updated with the current political scenarios, but still prefers reading only entertainment news on his apps! A female, in the age range of 40-50 can have immense interest in Sports! In this particular case, Collaborative Approach will fail because it might match preferences of females in the age range of 40-50 and come up with suggestions for Recipes, Movies, and Spiritual Guides etc. Different researchers have proposed a number of interesting theories regarding various approaches for analyzing user preferences. An interesting study in MIT concludes that the instead of asking users to rate the items on a scale of 0-5 as most of the Recommendation Systems do, users should be asked to rate the products in pairs. For e.g. a user who buys a laptop might also look for a laptop table in the same search. So, considering the ratings of a user for the laptop as well as the table suitable for the same, would give a more genuine result [12]. Privacy can also be an important factor to be considered while constructing user profiles. Although, most

of the times these surveys are anonymous, there is a huge possibility that the data from one set matches with data from another set and thus, the user identity can no longer be guaranteed to be safe.



Also, Recommending various items/products solely based on the user activities can be a total failure. Social Networking Site like Facebook shows a list of Recommended Pages to the user based on the activities of other users present in the Friend List of the user. The figure below shows Recommended Pages for Travelling, Movies, and TV Shows etc. Friends on Facebook are not necessarily friends with similar likes or dislikes. They can be acquaintances, colleagues, distant relatives and sometimes totally unrelated people. Recommended pages based on the data of such users can be totally futile, as the likes/dislikes might not necessarily match. Consider the case of the user in the figure below. The user never plays any games on Facebook and thus the suggestion of a game would be of no interest to the user. Also, the suggestion for “Learning Hyperdrive Inc.” might not be as appealing as the user probably would fail to understand the significance of what it means.

 **Shweta** | [Home](#) |  | 


 **Amit Arora's** birthday is today



 [Create Event](#)



Games You May Like



 **Hoop De Loop Saga**
Swathi Medavaram plays Hoop De Loop Saga.
 [Play Now](#)

Recommended Pages [See All](#)

 **Bay Area Travels and Tours, Inc.**
Shambhavi Bhorkar and Mrugan Bhorkar like this.
 [Like](#)

 **Learning Hyperdrive Inc.**
Abhay Bhorkar and 2 other friends like this.
 [Like](#)

 **Full House**
Sonal Behere-Goel and 11 other friends like this.
 [Like](#)

 **Life in pictures**
Suraj Mody and Sagar Shah like this.
 [Like](#)

Facebook © 2013
[English \(US\)](#) · [Privacy](#) · [Terms](#) · [Cookies](#) · [More](#)

Figure 8.Recommendations

Since this application is personalized and for registered users, the users would be aware that their reading patterns are being recorded. Since the emphasis is providing news articles as per the past preferences, secure information such as financial transactions, passwords etc. are not tracked, ensuring privacy of the users is respected.

4.2 Experiments:

As a part of the experiments, I created a user questionnaire to analyze the inclination of the users for the algorithms popularly used, viz: Collaborative Filtering and Content-Based Filtering.

I have considered a sample space of ten users with different social & educational backgrounds to conduct the experiments of the news recommendation application. I considered eight different news categories viz. Politics, Business/Finance, Sports, Entertainment, Technology, Education, Weather and World to study the behavior of users and stored the count as each user clicks any news article amongst the categories mentioned before. Content-based recommendation system would learn about user preferences based on the count of user clicks in a particular news category and would suggest the news in the similar categories. It can also recommend the articles based on the ratings that user has awarded while reading from a particular category earlier. Thus, content-based recommendation system follows a completely individualistic approach, where it tries to learn user preferences in various formats and continue recommending the news articles from preferred categories for a specific user. The graph below displays the recommendation results rankings per user in a pure content-based recommendation system.

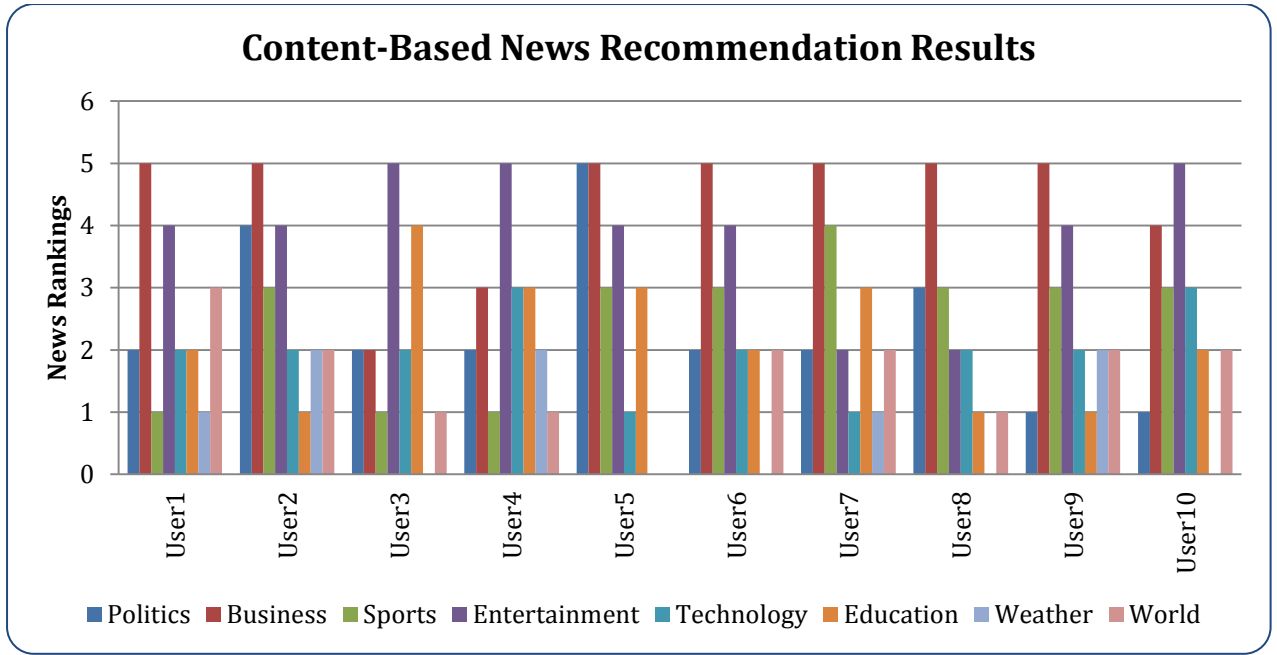


Figure 9. Content-Based News Recommendation Results

On the contrary, in case of a collaborative system, users would get news recommendations based on the user preferences that have similar user profiles or similar social, educational backgrounds or having similar hobbies, interests and inclinations. I have considered a subset of five users viz. User1 through User5 to have similar social, educational backgrounds as well as similar interests and hobbies. With this assumption, collaborative recommendation system would recommend news articles from the categories that have been visited most frequently, upon learning user preferences from such close-group users. Collaborative recommendation system would store the count of user visits per news category and calculate the average count of number of user visits in a particular category. With average count calculated, it would start recommending more news articles from that particular news category for all close-group users who have number of clicks less than the calculated average count. This will turn around the news recommendation result rankings observed in content based recommendation system. The graph below displays the recommendation result rankings for a close-user group in a purely collaborative recommendation system.

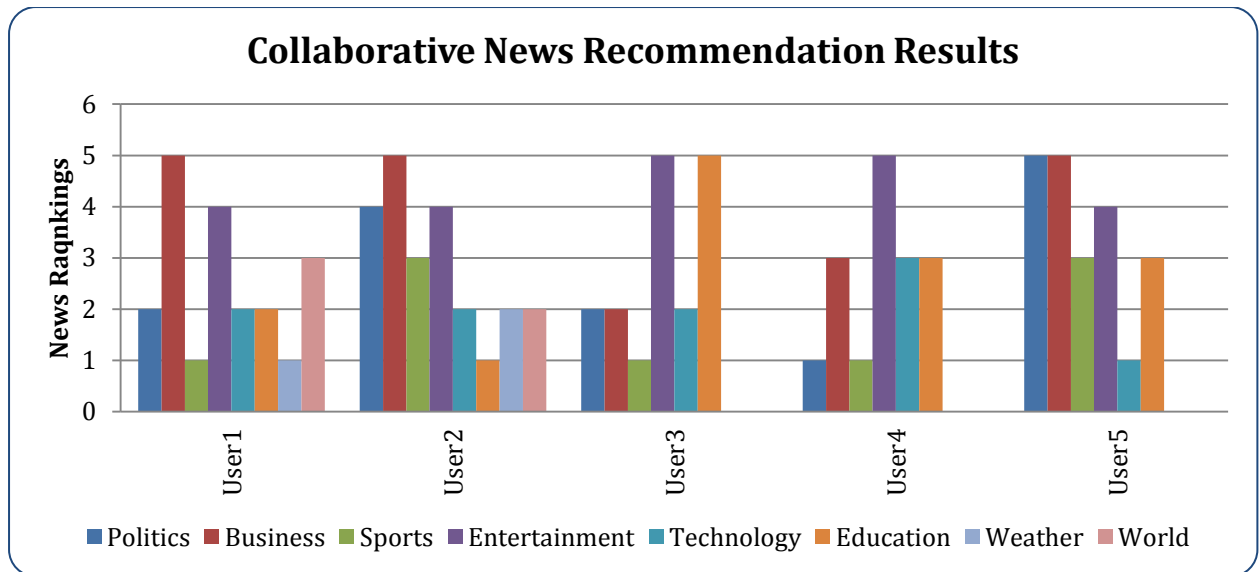


Figure 10. Collaborative News Recommendation Results

The following figures show the comparison between content-based and collaborative approaches

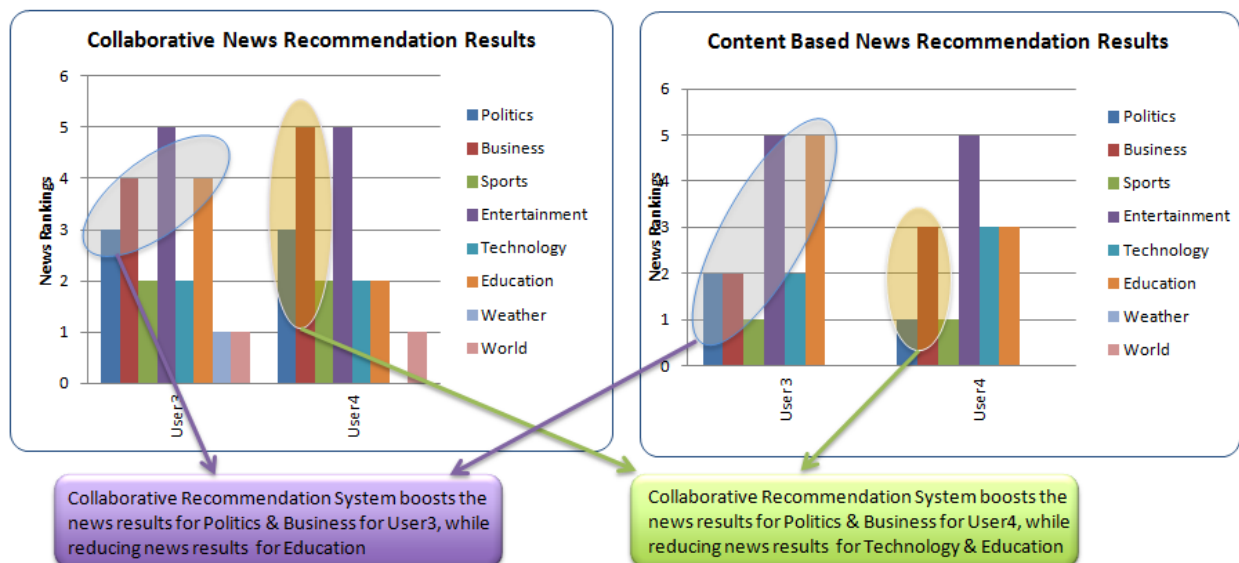


Figure 11. Comparison Results

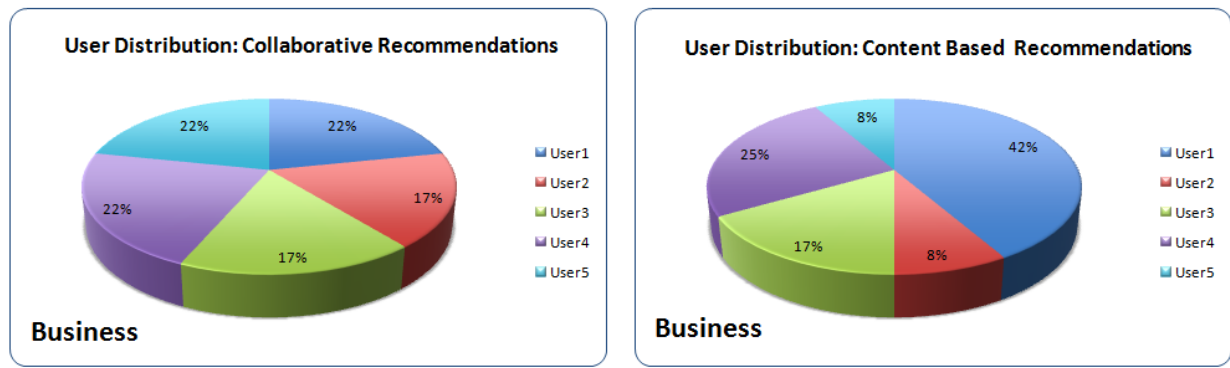


Figure 12. User Distribution Comparison

Based on the actual ratings, predictive ratings were created. For instance, if the rating for a particular category was high, the predicted ratings were unchanged. However, if a rating for a user for a particular category was less, the ratings were boosted using the Collaborative approach, based on the ratings of a group of users. The figure below shows the Hybrid Model and depicts a comparison between the three approaches.

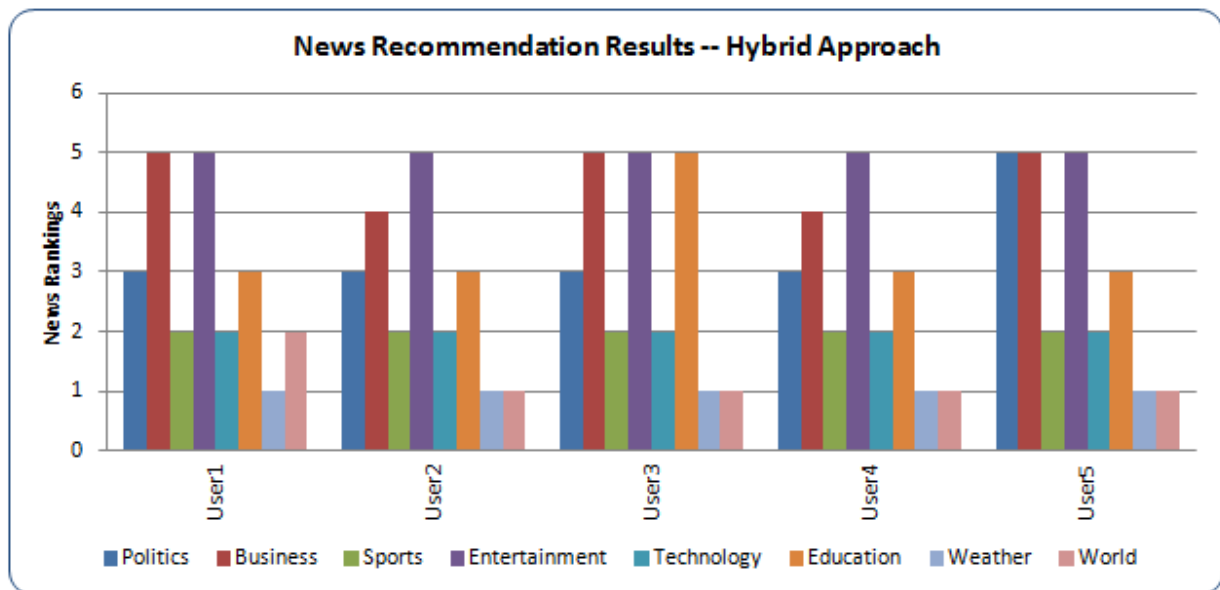


Figure 13. Hybrid Approach Results

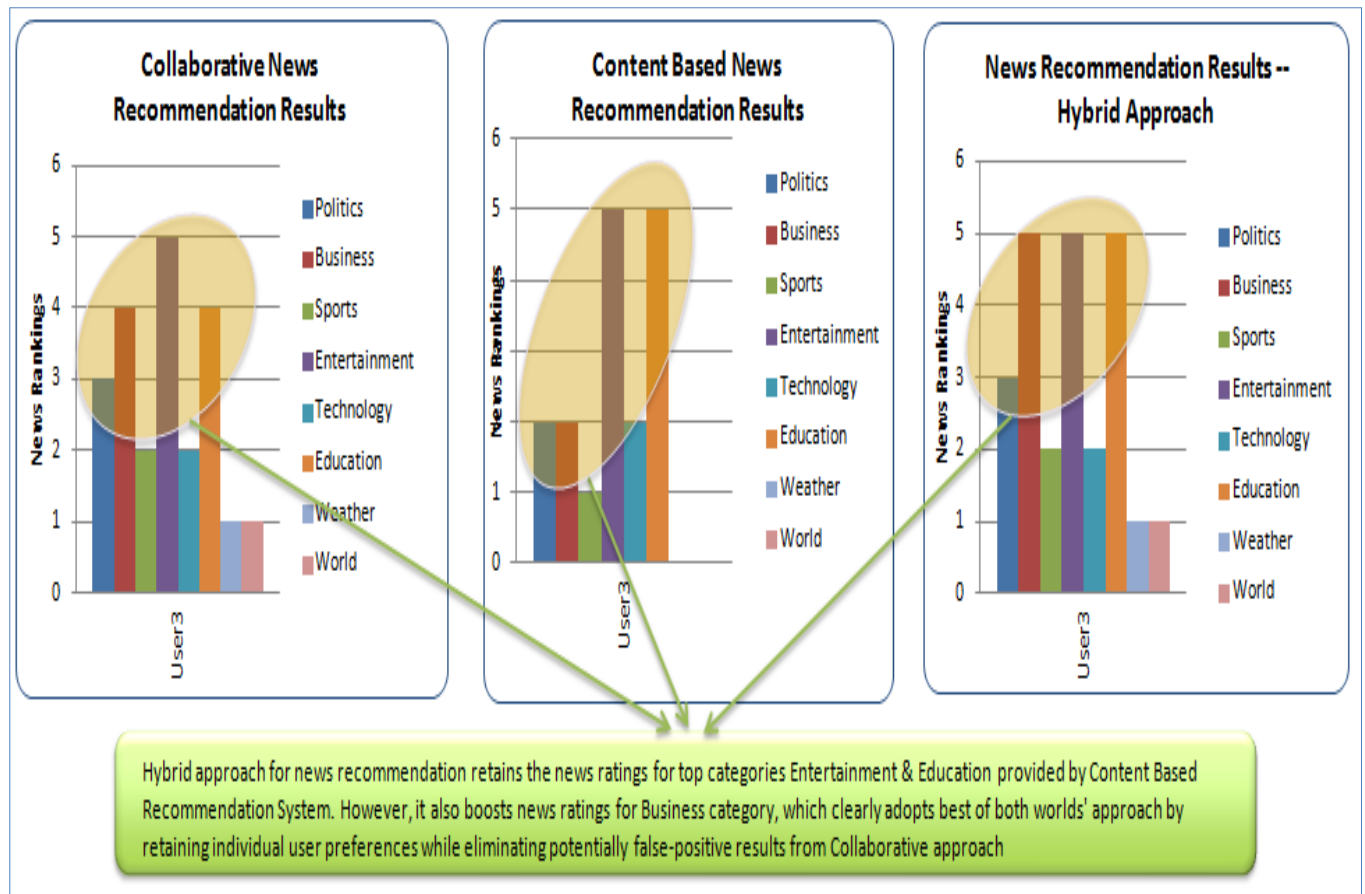


Figure 14. Hybrid Approach v/s Other Approaches

The survey consisted of 10 questions and was carried out by asking 10 users to answer the questions subjectively. Some of the questions were specific to the newsreader application, while a few others targeted at understanding, which Recommendation Algorithm worked better. The survey gave the following results:

- 1) Which newspapers do you like to read?

The most popular newspapers were BBC, Yahoo, NDTV, Times of India, CNN

- 2) What matters to you: Criticality of News or Popularity of the news items

Many a times, critical news are popular news. Most of the users liked to browse through the breaking news and liked to read in detail the articles of their personal interest

- 3) Do you prefer to read newspapers which you have already been reading or you are open to read new news articles, trending articles

Some of the users liked to read only trending articles, while most of them liked reading from their favorite categories. They were open to reading new articles, but the frequency was intermittent.

4) Which categories do you like to read?

Various categories like Business, Sports, Politics, World News were named

5) You are an Economist; do you like to read business news more than any other news section?

For this question, we divided the users in different groups, based on their Social Status, Occupation, Gender, and Age. The answers we received were surprising! The normal notions for tying preferences to the attributes, which normally define the likes/dislikes of people, were defied.

6) Do you like recommendations in form of suggestions appearing on the screen when you access a particular website or smart-phone application

This question could not elicit a strongly positive or negative response. Most of the users said that they like to see recommendations only when they are in real need of any suggestions!

7) If Yes, what kind of recommendations do you like?

Considering that not all the users had a technical background, instead of telling the users about the algorithms used, they were asked if they liked suggestions based on the items they were looking for (Content-Based Filtering Technique or Item-Based Collaborative Technique) or if they looked for suggestions based on what other users preferred, trending items and so on (User-Based Collaborative Technique). Users were able to answer in terms of the technique they preferred based on the items under consideration.

- 8) Do you rely on reviews, ratings given on the websites before reading anything, buying anything

It was found that the users found ratings and reviews for e-commerce sites, apartments etc. important. However, some of the users specifically mentioned that they could not rely on these reviews and ratings completely and had to go through the details thoroughly before making any decisions based on them. For eg: Some reviewers praised about a certain property, however the rating given was only 3/5 and there was no clear explanation why the rating was average.

- 9) Do you thoroughly read news articles in detail or just browse through?

Most of the users answered that they just browsed through the various applications on the smartphone, but in case of few articles, they bookmarked them for reading thoroughly.

- 10) How did you like the newsreader application?

The users liked the simple interface provided by the application. They liked that they could read the articles from the categories, which they preferred, also they liked that the application provided links for various other articles too. They were happy that their privacy was not compromised to elicit recommendations.

5.Conclusion and Future Work

This paper studies the Content-Based and Collaborative-Based Approaches in detail and comes up with a Hybrid Model for Recommendations. The Hybrid Model wins over the individual approaches in terms of encompassing positive aspects of both the algorithms and tries to remove the shortcomings. Since this is a hybrid model, even when the actual recommendations start after a week's activity, the user still gets to read popular newspapers. Different attributes essential for the success of a Recommendation System such as Novelty, Trust, Coverage, and Privacy are addressed. New features can be added to the application, which give the user a facility to add categories or newspapers. Cloud based approach for storing data can be used to improve the overall storage and processing of data.

6. List of References

- [1] Wikipedia Information on News
<http://en.wikipedia.org/wiki/News>[Retrieved on March 2013]
- [2] Android Apps Data
<http://android.appstorm.net/roundups/lifestyle-roundups/35-fantastic-news-apps-for-android/>[Retrieved on March 2013]
- [3] Martin P. Robillard, Robert J. Walker, Thomas Zimmermann., Recommendation Systems for Software Engineering. In *IEEE Computer Society*
- [4] Recommendation Systems
http://en.wikipedia.org/wiki/Recommender_system [Retrieved on March 2013]
- [5] Xiaoyuan Su and Taghi M. Khoshgoftaar., A Survey of Collaborative Filtering Techniques. In *Advances in Artificial Intelligence, Article ID 421425*
- [6] Hans van der Heijden, Gabriele Kotsis, Reinhard Kronsteiner2., Mobile recommendation systems for decision making ‘on the go’. In *Proceedings of the International Conference on Mobile Business (ICMB’05)*
- [7] Sean B. Minkel., Technology Review [Retrieved on March 2013]
- [8] Robin van Meteren, Maarten van Someren., Using Content-Based Filtering for Recommendation
http://www.ics.forth.gr/~potamias/mlnia/paper_6.pdf
- [9] Marko Balabanovic, Yoav Shoham., Content-Based, Collaborative Recommendation
http://sentra.ischool.utexas.edu/~i385q/spring2005/readings/Balabanovic_Shoham-1997-Fab.pdf
- [10] HengSong Tan, HongWu Ye., A Collaborative Filtering Recommendation Algorithm Based On Item Classification
- [11] Kam Fung Yeung, Yanyan Yang., A Proactive Personalized Mobile News Recommendation System
- [12] Improving recommendation systems
<http://web.mit.edu/newsoffice/2011/compare-recommendation-systems-0708.html>
[Retrieved on March 2013]
- [13] How Foursquare Works
<http://computer.howstuffworks.com/internet/social-networking/networks/foursquare2.html>
[Retrieved on March 2013]

- [14] Feng Gao, Yuhong LI, Li Han, Jian MA., InfoSlim: An Ontology-Content Based Personalized Mobile News Recommendation System. In *5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009.WiCom '09*
- [15] Mukund Deshpande and George Karypis, University of Minnesota., Item-Based top-N Recommendation Algorithms. In *ACM Transactions on Information Systems*, Vol. 22, No.1, January 2004
- [16] Prem Melville and Raymond J.Mooney and Ramadass Nagarajan., Content-Boosted Collaborative Filtering for Improved Recommendations. In *Proceedings if the Eighteenth National Conference on Artificial Intelligence(AAAI-2002),pp.187-189, Edmonton, Canada, July 2002*
- [17] Sandra Garcia Ezparza, Michael P. O'Mahony, Barry Smith., A Multi-Criteria Evaluation of a User-Generated Content Based Recommender System
- [18] Buckley, C., and Salton, G. Optimization of relevance feedback weights. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (Seattle, July 1995)*
- [19] Guy Shahani, and Asela Gunawardans., Evaluating Recommendation Systems

7. Appendix

The Dataset used has been attached.



Data.xlsx