Purakala
(UGC Care Journal)
ISSN:0971-2143
Vol-31-Issue-31-May-2020

Engineers Guide: Recommendation System

[1]Jayashree Hajgude, [2]Yash Bhardwaj, [2]Rohit Kataria, [2]Yogesh Lulla
[1]Assistant Professor, [2]Student
Department of Information Technology
Vivekanand Education Society's Institute of Technology
Mumbai, Maharashtra, India

Abstract: Several studies have reported the importance of co-curricular and extra co-curricular activities in college along with a good academic score; however, reprogramming this into the current Education system is yet to be accomplished. The paper describes an overview of the website which helps the end users in the above mentioned problem. Firstly, a recommendation engine enables the user to rate a particular book and get recommendation according to his rating. Secondly, along with the recommendation the paper also provides an event forum displaying all the co-curricular events in the college. Finally, using web crawling and scraping techniques web content of different hackathon events from different sites is obtained.

Keywords: Collaborative filtering, Content based filtering, Recommendation system, Scraping.

1. INTRODUCTION

Recommendation system for the books will be implemented as books play the most important role in any student life. Strong academic score cannot be achieved without books. Books help the students to understand the concept in the best possible way. Many of them have their own recommendation system to recommend books to the buyers. This paper presents a new approach for recommending books to the buyers. This system combines the features of content filtering and collaborative filtering to produce efficient and effective recommendations. Extra curricular activities are important for everyone with respect to grade or standard. Extra curricular activities just don't improve students skills but also develop their brain. It is found that students who performed different extracurricular activities were a bit smarter and had overall developing skills. Also, there is a chatbox and forum section using which customers can communicate with each other and resolve their queries. Also, there is a scraping section where users will get to know about upcoming hackathons along with necessary details.

1. **Rating based input**: Different users of the website will rate the books according to their liking and depending upon that, new users will get suggestions of those books.
2. **Content based input**: At the time of signup, the users have to enter their details along with their branch and depending upon that also, the users will get suggestions.

2. Content Based Recommendation

Content based recommendation is the most common method used for recommendation. The basic operations performed by a content based recommendation system consists of various factors like matching user various data like location, age, gender, branch and rated items list on the site stored in his account with similar items having common specifications, in order to recommend new items meet the users interests. Content based recommendation is usually based on two approaches that is Analysing the description of the items and Building users profile and item profile from user rated content.[2]

1. **Analysing the description of the content :** In this approach, the system will suggest anything similar which the users had liked before. The content of the books which the user has gone through, will be compared with all the remaining books which are present in the database and the books which matches successfully will be recommended to the user.

ISSN:0971-2143
Vol-31-Issue-31-May-2020

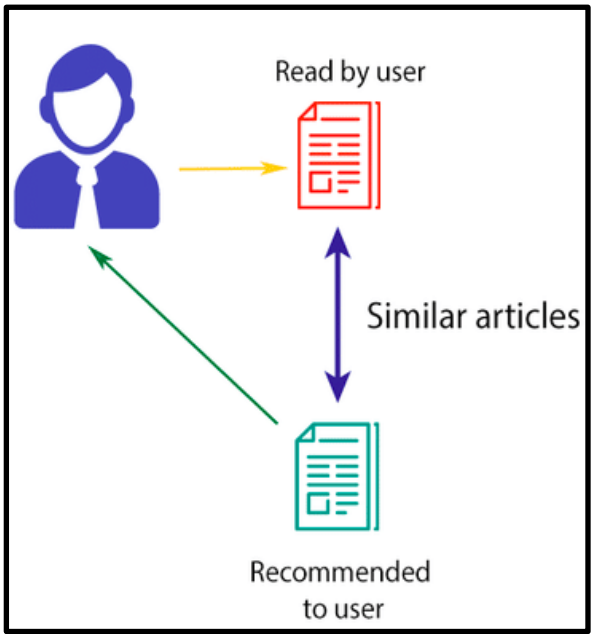

figure 1: content based filtering

2. **Building user profile and item profile from user rated content** : This model constructs user profile based on user preferences and tries to generalize data. Normally, generalisation is done through machine learning techniques, which are able to understand examples of user interests, starting from the books which are liked or disliked in the past. In case of books, the category of books as well as raw text is used to create users profile and books profiles.

This method is suitable in the situations where the description of the items are known but not of the users. In Content based recommendation, the keywords are extracted from those items and then items will be identified with those keywords and further users' profiles will be created to indicate the different types of items users like. In other words, content based recommendations try to recommend items that are similar to those that the users have liked in the past.[1][3]

3. Collaborative Filtering

Collaborative filtering is the most common approach used for recommendation. In order to find out the quality of books, collaborative filtering is a must. Collaborative filtering works on historical preferences of the users on a particular set of items. Collaborative filtering is the most powerful and traditional way of recommendation. In Collaborative filtering, it is assumed that the users who have agreed in the past will also agree in the future. Depending on the users preferences, it is normally expressed as two categories :

1. **Explicit Rating:** Explicit Rating is a rating which is directly given by the users to a particular set of books from the range of 10. From the Explicit Rating, direct feedback of that user for that particular book is known.
2. **Implicit Rating:** Implicit Rating is a rating which is indirectly given by the users. Implicit Ratings includes page views, number of clicks, purchase records etc.

The basic idea behind collaborative filtering is that it works in collaboration with other users. For instance, there are three users i.e. user 1, user 2 and user 3. If user 1 likes books A, B, C and user 2 like books B, C, D and user 3 like books B, C, E. Since books B, C is similar for all the three users and therefore the new book which user A will like in the future, will be recommended to user 2 and 3 also and so on.[1]

ISSN:0971-2143
Vol-31-Issue-31-May-2020

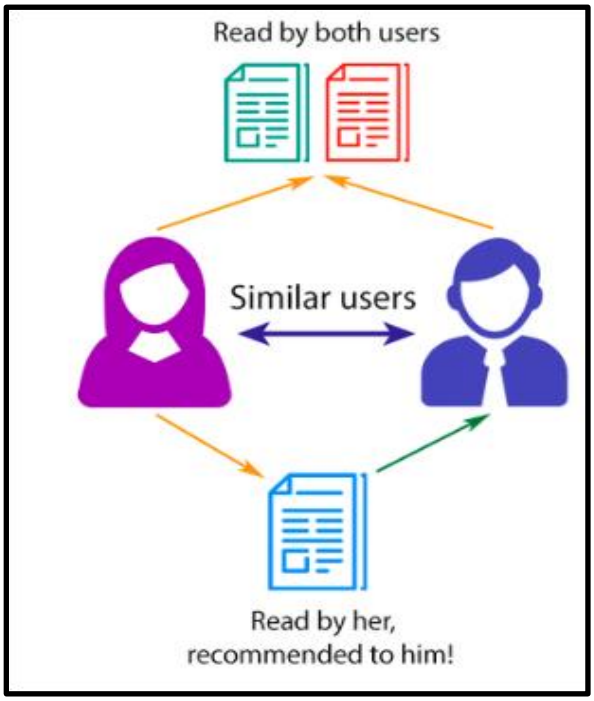

figure 2: collaborative filtering

3.1 Finding out similarities between books:

There are a number of different algorithms to find similarities between books like Cosine similarity, Pearson correlation, Euclidean, Manhattan and so on. This paper uses cosine similarity to find matching books.[1]

3.2 Cosine similarity:

The dot product between two vectors is equal to projection of one of them on the other. Therefore, dot product between two same vectors is equal to their squared module and if two vectors are perpendicular then the dot product will be zero. For n dimensional vectors, the dot product can be calculated as follows

$$\boldsymbol{u} \cdot \boldsymbol{v} = [u_1\ u_2\ \dots\ u_n] \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n = \sum_{i=1}^{n} u_i v_i$$

Dot product.

figure 3: dot product

The dot product is important while defining similarity between products as it is directly connected to it. We can define similarity between two vectors u and v as the ratio between their dot product and the product of their magnitudes.

$$similarity = cos(\theta) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \|\boldsymbol{v}\|} = \frac{\sum_{i=1}^{n} u_i v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \sqrt{\sum_{i=1}^{n} v_i^2}}$$

Similarity.

figure 4: cosine similarity

ISSN:0971-2143
Vol-31-Issue-31-May-2020

Using the above equations, similarities between all the users will be calculated. The value will be in the range of 0 and 1. Depending upon the values, the users will be considered either similar or dissimilar and therefore the recommendation of the similar users will be shown accordingly and the users which are not similar to any of the other registered users, the books will not be recommended to them.[2]

4. Chat Box and Chat Forum

In the chat box section, users can communicate with other users and in chat forum sections users can post their doubts, queries etc and that can be viewed and answered by all other users. All messages of the chat box and all queries of the chat forum will be stored in the database along with a datetime stamp. These messages load from the database when users click on the username located at the left of the screen.

5. Web Scraping

Web scraping also called Web harvesting or Web extraction is a process of extracting data from websites. Web scraping software uses HTTP (Hypertext Transfer Protocol) to scrape data. The system has two parts namely fetching and extracting.

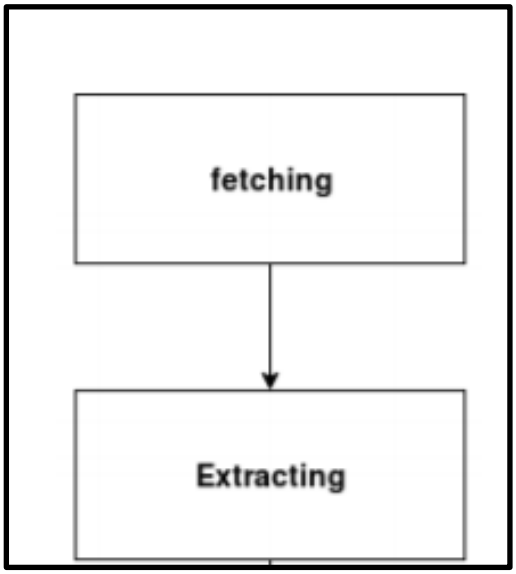

figure 5: stages in web scraping

5.1 Fetching:

The second part is the extraction of the data, in this part the Visual basic module does the job of getting the relevant data from the system. In this architecture, the web scraping is dichotomized half on the online proces half on the online process and half on the offline process.

Here online process means that you need an internet connection with proper bandwidth and further offline tools are used to extract information from the data that is received after exploring the source code . Firstly, JAVA application is used to extract source code from the website. Secondly, visual basic macros are being used to extract relevant information from the source code and further this data is shown to the user in a structured form. This data can be used for contact making and lead generation. Further databases can be formed on this basis. This data can be used for demographic analysis, resource analysis etc. The working of a web scraper is quite simple, it starts with a list of URLs to visit which are called seeds. The web scraper will try to scrape data but yet there is some website that does not allow scraping which makes it impossible to scrape them.[6] The web scraper then fetches the source code of all the web pages given in the list only of websites which allow scraping. From this source code the main process of extracting will start afterwards.[4]

ISSN:0971-2143
Vol-31-Issue-31-May-2020

5.2 Jsoup:

Jsoup is an open source java library which consists of different methods which are used to extract and manipulate HTML web pages. The following things can be done using Jsoup:

1. Scrape and parse HTML from a file, URL or string.
2. The data can be found and extracted using DOM traversal or CSS selectors.[5]
3. Different things can be manipulated like text, HTML elements and attributes.
4. XSS attacks can also be prevented.

The paper represents the basic idea for scraping using jsoup. The structure for some fixed set of websites which displays information about upcoming hackathons is studied and depending upon that, code is developed using jsoup to scrape the data about various upcoming hackathons along with the necessary details and display that data on the website.[7]

6. Results

Books Recommender System: Books of a similar genre and storyline-up are listed based on the previous books read by the user. The goal of the most recommendation system is to predict the buyer's interest and recommend the books accordingly. This book recommendation has considered many parameters like content of the book and quality of the book by doing collaborative filtering of ratings by the other buyers. This recommender system also uses an associative model to give stronger recommendations.

Chat Forum: A chat forum is also provided where a user can post his queries and can get appropriate answers. The chat forum acts as an open question answer platform where the answer gets timestamped with the details of the user. The user can either view the answer or give the answer.

Chat Box: Users can chat with each other in the chat box , to have personalized communication with different registered users. Having personalized communication with colleagues of the same field makes the feature more appropriate in this context.

Web Scraping: With the help of Web scraping , data about upcoming hackathons will be scraped from different websites and displayed on the website along with other necessary details.

7. Conclusion

Providing the user with the best books will help them to score well in their academics**.** Participating in extracurricular activities sharpens the brain and also gives a strong experience which only curricular activities cannot provide. This paper would help the engineering students to participate in such activities of different colleges.

8. References

[1] Zhao kai ,Lu Peng-Yu ,“ Improved Collaborative filtering approach based on User similarity Combination”,IEEE International Conference on Management Science & Engineering ,Helsinki, Finland ,2014,pp 238-241.

[2] Pijitra Jomsri, “Book recommendation System for Digital Library based on User profile by using Association rule”, Thailand, IEEE, 2014, pp 130-134.

[3] Anand Shanker Tewari, Kumari Priyanka ,”Book Recommendation System based on Collaborative filtering and association rule mining for college students”,IEEE,2014,pp 135-138.

[4] Osmar Castrillo-Fernández, "Web Scraping: Applications and Tools", European Public Sector Information Platform Topic Report No. 2015 / 10, December 2015.

[5] Kanehisa M, Goto S, Sato Y, et al. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012;40:D109–14.

[6] Glez-Pen‹a et al. , "Web scraping technologies in an API world", Briefings in Bioinformatics Advance Access, doi:10.1093/bib/bbt026, published April 30, 2013.

[7] "Implementation of Web Crawler", Pooja gupta, Member, IEEE, Kalpana Johari, Member, IEEE.