A RECOMMENDER SYSTEM FOR ACADEMIC PROJECTS REPOSITRY

Bolanle F. Oladejo
oladejobola2002@gmail.com
Department of Computer Science, University of Ibadan, Ibadan, Nigeria.

Abstract—The Academic Project Recommender System is concerned with providing final year students of degree programs with a plethora of past project reports related to different fields of interest. In a local library, project reports are stored and basically classified according to the respective year of submission, and type of programs, as either postgraduate or undergraduate. Hence, it requires painstaking effort to search for past projects based on specific areas of research since project reportsare usually indexed on title and keywords. In order to solve the underlying problem of difficulty in accessing specific research areas from project reports, this workaimed at development of Academic Project Recommender System. Content based recommender and map reduce techniqueswere applied to project reports to facilitate retrieval of a recommended topic. Final year students could retrieve project reports through search termsrelated to particular field of interest.A Map Reduce technique is used to check for the frequency of occurrence of the keywords in each project report. Consequently, it indexes project reports according to frequency. This ensures that the search results are organized in the order of relevance to the user's search preference.The system promotes research continuity by providing students with several related projects through search and recommendationof project topics.

Index Terms—Final Year Project, Information retrieval, Map reduce technique, Content based Recommender, Recommended Project Topics

1 INTRODUCTION

Education is a social process - ,"a process of living and not a preparation for future living". [1]. [2] defines education as "the aggregate of all the processes by which a child or adult develops the abilities, attitudes and other forms of behavior which are of positive value to the society in which he lives, that is to say, it is a process of disseminating knowledge either to ensure social control or to guarantee rational direction of the society or both". It is the process of imparting or acquiring structured and organized information, and instructions. It has overtime become a necessity rather than a mere showcase of wealth and ostentation. Education could basically be either formal or informal. Informal education refers to the process of gaining knowledge without necessarily being within a normal school setting or environment. It does not follow any pre-specified curriculum. On the other hand, formal education usually follows a specific pedagogy. The basic pattern for students who receive formal education in Nigeria is pre-school, primary school, secondary and then tertiary (higher) [3]. Higher education usually involves working towards a degree-level or foundation degree qualification. It is typical for each of these educational stages to have a stipulated duration or at least a range of time in which that stage would be completed. After each stage, the students graduate to another level.

It is a norm for students in higher institutions to work on projects in their final year. This fact therefore implies that students would need to decide on the project topics to work on. In partial fulfilment of graduation requirements, final year (FY) students of tertiary institutions conduct research and carry out major projects. FY projects are very important to the success of students. It gives rise to maturity and a sense of responsibility because it prepares FY students to face problems that might come up in their future professional career [4]. FY projects help to synthesize all the knowledge and experience that the students have gained in the past, and further creates a platform for the students to do a lot of study, do more research and to solve a defined problem. It also helps the students to work within a limited time frame as well as with few resources.

It is therefore important for FY students to work on projects that are extremely beneficial. FY topics are searched for in a couple of ways. A good number of students use the internet as a source for project topics, others come up with exceptional project ideas in order to solve a very pressing societal problem or to meet the need of a particular sect of people, others are simply given project topics by their supervisor based on their area of specialization, some others check their schools library for past project recommendations in order to extend the projects done in the past. The sources of project topics are somewhat inexhaustible.

In a good number of tertiary institutions across the world, projects are not handled individually but as a synergy of efforts from a group of students under a particular supervisor. As a result of this, more often than not, the scope of topics online are usually wide and could barely be handled by just one individual. It is also usually difficult to use these project ideas because they sometimes do not relate to the situation of the school or environment in question. In a department where past project reports are not properly organized and indexed, searching for specific past project reports become a thorn in the flesh for the students. Due to this anomaly, a number of students find it quite burdensome to settle for project topics. Even when they eventually do, problems arise if they want to verify if their topic has been handled before and to what extent. Having access to the past project works of former students help the search for project topics much easier. This is because the recommendations made by the authors of the past projects could give students an insight on what project focus, or further work to existing project outcome they could engage in. Therefore, this work aimed at the design and implementation of an Academic Project Recommender System (APRS). The rest of this paper is divided into four sections. Section 2 considers the relevant theoretical background and review of related works. Section 3 focuses on modeling of the Academic Project Recommender System presented in this work while discussion ofimplementation and results is considered in section 4. The paper is concluded in section 5 with further work's recommendation.

2 THEORETICAL BACKGROUND

The background of the study and review of related works are discussed in the section.

2.1 Information Retrieval

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers). The field of information retrieval also covers supporting users in browsing or filtering document collections or further processing a set of retrieved documents. It is often approached by first manually classifying some documents and then hoping to be able to classify new documents automatically.

Information retrieval systems can also be distinguished by the scale at which they operate, and it is useful to distinguish three prominent scales. In web search, the system has to provide search over billions of documents stored on millions of computers. Distinctive issues include the need to gather documents for indexing, being able to build systems that work efficiently at this enormous scale, and handling particular aspects of the web, such as the exploitation of hypertext and not being fooled by site providers manipulating page content in an attempt to boost their search engine rankings, given the commercial importance of the web. The other extreme is personal information retrieval. In the last few years, consumer operating systems have integrated information retrieval. Email programs usually not only provide search but also text classification: they at least provide a spam (junk mail) filter, and commonly also provide either manual or automatic means for classifying mail so that it can be placed directly into particular folders. Distinctive issues here include handling the broad range of document types on a typical personal computer, and making the search system maintenance free and sufficiently lightweight in terms of startup, processing, and disk space usage that it can run on one machine without annoying its owner. In between is the space of enterprise, institutional, and domain-specific search, where retrieval might be provided for collections such as a corporation's internal documents, a database of patents, or research articles. In this case, the documents will typically be stored on centralized file systems and one or a handful of dedicated machines will provide search over the collection.

However, the major problem of information retrieval is the difficulty in determining the relevance of query results to users' need [5].

2.1.1 Information Filtering

Information filtering systems are designed for unstructured or semi-structured data, as opposed to database applications, which use very structured data. The systems also deal primarily with textual information, but they may also entail images, voice, video or other data types that are part of multimedia information systems. Information filtering systems also involve a large amount of data and streams of incoming data, whether broadcast from a remote source or sent directly by other sources. Filtering is based on

descriptions of individual or group information preferences, or profiles that typically represent long-term interests. Filtering also implies removal of data from an incoming stream rather than finding data in the stream; users see only the data that is extracted [6].

Information filtering is a term used to describe a variety of processes involving the delivery of information to people who need it. Although this term is appearing quite often in popular and technical articles describing applications such as electronic mail, multimedia distributed systems, and electronic office documents, the distinction between filtering and related processes such as retrieval, routing, categorization, and extraction is often not clear. It is only by making that distinction, however, that the specific research issues associated with filtering can be identified and addressed.

Information filtering systems deal primarily with textual information. In fact, unstructured data is often used as a synonym for textual data. It is, however, more general than that and should include other types of data such as images, voice, and video that are part of multimedia information systems. None of these data types are handled well by conventional database systems, and all have meanings that are difficult to represent.

2.1.2 Map Reduce (MR) Search Technique

Map Reduce is a search technique which takes fragments of information about an object scattered through a large input file, and collecting them such that they are next to one another in the output [7]. A map reduce variant could be used to search for data in a large collection such that the keywords' frequencies are checked and the results are ranked according to the highest frequency.

2.2 Recommender Systems (RS)

A Recommender systems is a subclass of the information filtering system. A recommender system or recommendation system is an extension of web applications that predict user responses to options. It makes use of user preferences and priorities. Recommender systems have become extremely common in recent years, and are applied in a variety of applications. Recommender systems typically produce a list of recommendations in a number of ways.

Recommender systems are a useful alternative to search algorithms since it helps users discover items they might not have found by themselves. Interestingly enough, recommender systems are often implemented using search engines indexing non-traditional data.The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them.

There are several techniques and approaches to recommender system which include content based recommender system, collaborative recommender system and hybrid recommender system.

2.2.1 Content Based RS

The system learns to recommend items that are similar to the ones that the user liked in the past. The similarity of items is calculated based on the features associated with the compared items [8]. For example, if a user has positively rated a movie that belongs to the comedy genre, then the system can learn to recommend other movies from this genre [9].

Content-based filtering exploits the content of data items to predict its relevance based on the user's profile as shown in figure 1. Research on content-based recommender systems takes place at the intersection of many Computer Science sub-fields, especially Information Retrieval and Artificial Intelligence.

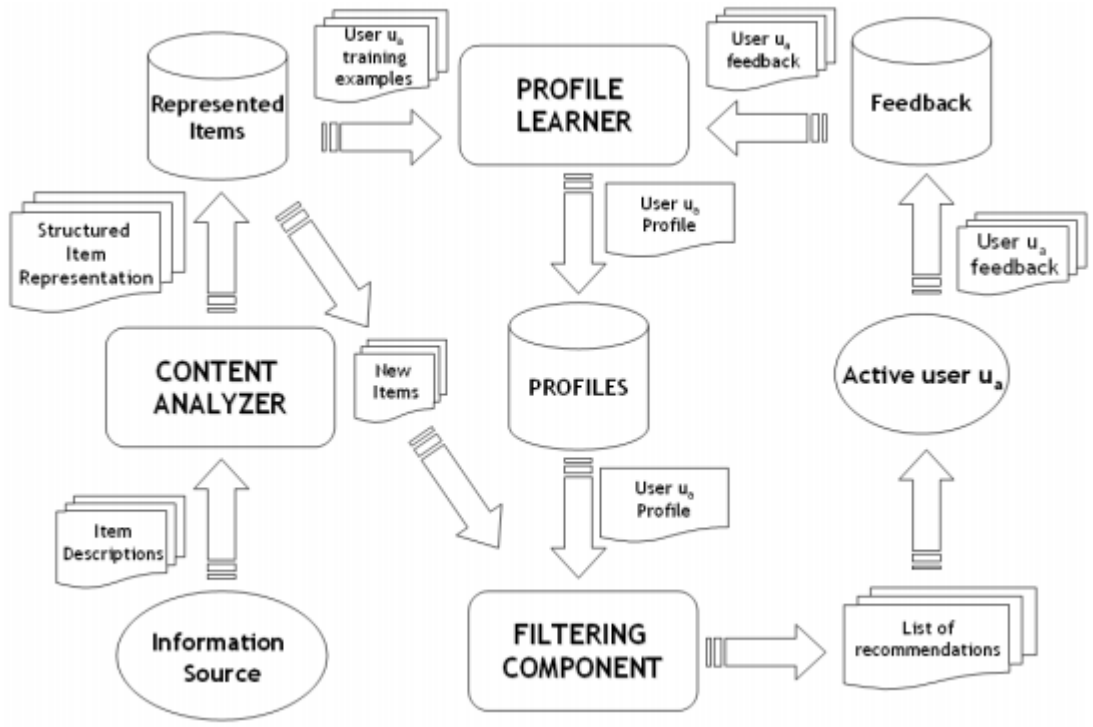

Fig.1. High level architecture of a Content-based Recommender [10] (Lops, Gemmis, &Semeraro, 2012)

2.2.2 Collaborative Filtering RS

Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users [9]. (Francesco, Rokach, & Shapira, 2011).

A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems [11]. (Millan, Trujillo, & Ortiz, 2007).

Collaborative Filtering is based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past.

Collaborative filtering approaches often suffer from three problems: cold start, scalability, and sparsity [12]. In the case ofCold Start,Collaborative recommender systems often need a large amount of existing data about a user before it is able to make accurate recommendations. In many of the environments in which these systems make recommendations, there are a large number of users. Thus, a large amount of computation power is often necessary to calculate recommendations, resulting to Scalabilityconstraint. Also, in collaborative filtering systems, users are typically represented by the items they have purchased or rated. Dueto Sparsity, it is highly probable that the similarity (or correlation) between two given users is zero, rendering collaborative filtering useless as the most active users will only have rated a small subset of the overall database.

2.2.3 Hybrid RS

Hybrid RSs are based on the combination of the above mentioned techniques. Research has shown that a combination of both approaches could amount to more effectiveness of the recommenders system. The hybrid approach could function in two ways. The two approaches could be combined to produce one major model which is treated as a new technique. Also, each approach, that is, the collaborative or content based can be used separately, but on the same project. The result produced by hybrid recommender is less prone to irregularities [13].

2.3 Academic Projects

Projects are undertaken by students in tertiary institutions in partial fulfilment of graduation

requirements. These projects are set as a visible evidence that the theories and fundamentals taught by lecturers are properly imbibed by the students. It allows students to develop problem solving, analysis, synthesis and evaluation skills.Final year projects fulfils a purpose of synthesizing all the knowledge gained throughout the early years in the university vicinity in order to solve a particular problem. In this salient process, students are moved to apply all knowledge gained. This allows for maturity on the part of the students [14]. (López, 2003).

As a Final Year Project is the last educational activity to be fulfilled before getting a degree, it prepares students for the outside world. It exposes them to the likely problems that could be faced beyond school. It also further builds the level of independence of students since the projects are usually undertaken individually or in groups of little numbers.

Students are assigned to several supervisors whose basic responsibility is to advice, assist, guide, and encourage students on their projects.With the wealth of knowledge possessed by a supervisor, the difficulties encountered by students would be attended to and resolved. By such means, a teacher contributes in facilitating the learning process, and at the same time stimulates students to progress little by little in this activity.

2.4 Review of Related Works

The review of some related works are presented subsequently.

2.4.1 A Library Recommender System based on Content Based Filtering Using the K-Means Algorithm

The work by [15]focuses on the development of a system which could reduce the workload and time of analyzing book borrowing transactions with respect to books borrowed, books that are always on demand or most preferred books. The system could adapt to the borrowing trends and the changing demands and preferences of the library user or borrower recommend additional books. The work demonstrated that content based filtering algorithms could be used to make efficient recommendations based on items analysis and implicit user preferences like other collaborative algorithms.

2.4.2 Matchbox Recommender System

One of the great challenges in the interconnected world of the web is to connect people with other people, content, or products they care about. This challenge appears in tasks like product recommendation, social matchmaking, targeted advertising and content filtering. Interestingly, the problem also occurs in more technical challenges such as picking the right algorithm for a given problem. The work by [16] focuses on application of Matchbox whichis based on Bayesian technique to address the stated challenge. Matchbox has the ability to learn about people's preferences from observing how they rate items such as movies, content, or other products. Also, based on observations Matchbox, is then able to recommend new items to users upon request.Matchbox has been designed to use the available data for each user as efficiently as possible. Its learning algorithm is designed specifically for the large streams of data typical for web-scale applications.

Its main feature is that Matchbox takes advantage of meta-data available for both users and items. This means that things learnt about one user or item can be transferred across to other users or items.

2.4.3 Research Repository Management System

In [17], Research Repository Management System (RRMS) was developed to curb the challenge of lost or misplacement of past academic projects reports. K-Means algorithm was used to classify relevant projects' features based on area of specialization to aid efficient retrieval of useful information. The similarity between RRMS and this work is the focus on academic projects while the discrepancy centers on the techniques of information retrieval for guiding decision of prospective project students.

3 RESEARCH METHODOLOGY

In a local library, project reports are shelved and categorized as either undergraduate or postgraduate. However, within each classification, the project reports are not catalogued and hence it is difficult to search through. This work thus addressed this constraint.

A content based recommending technique and a map reduce variant was adopted in the Academic Project Recommender System (APRS). This system includes the development of a repository of project contents for recommendation purpose.

3.1 Architecture of Academic Project Recommender System

The architecture of Academic Project Recommender System (APRS) depicted in figure 2 is the overall structure of the system with emphasis on interaction existing among several core users of the system on a network.The users (Admin, Graduate Student and Final Year students) associate with the core functions of the system via an institution network.For instance, a FY student at the completion of his study uploads a copy of project report, which the Admin verifies before updating projects repository. A prospective FY project student explores to obtain recommendation for project topics.

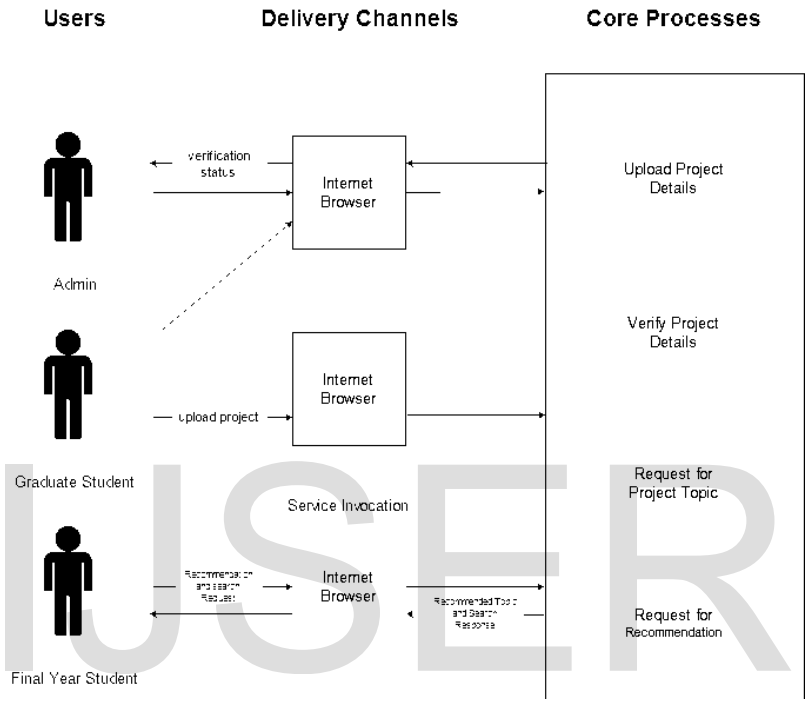

Fig. 2. Architectural design of Academic Project Recommender System

3.2 Relational Data Model of Academic Project Recommender System

Figure 3 depicts the logical data model for repository of academic projects and details of a student's search and recommendation of Academic Project Recommender System (APRS).

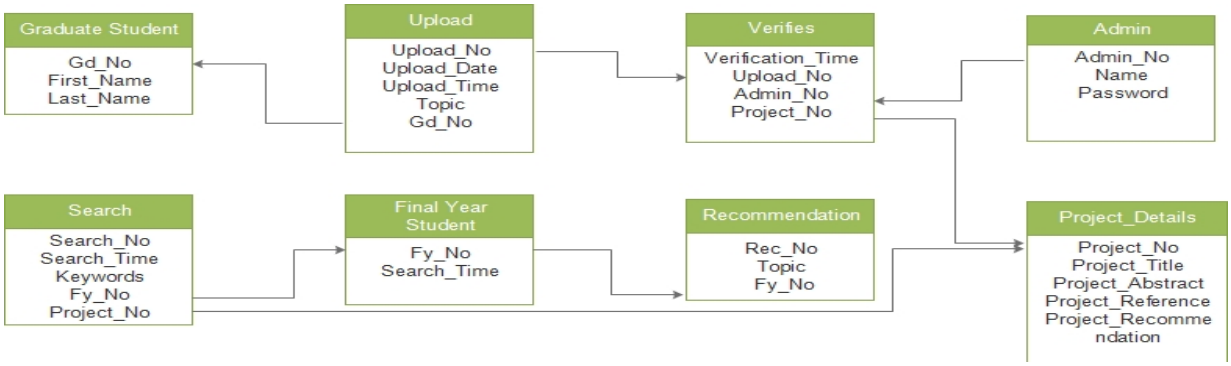

Fig. 3. Logical Data Model of Academic Project Recommender System

3.3 Logic Modeling of Academic Project Recommender System

For the academic project recommender system, two major techniques were employed.The map reduce variant techniquewas used in the search process and the content based recommendation technique was applied in the recommendation process. Figure 4 depicts the algorithm for the search process.

```
Start
Step 1: Accept search input from the user
Step 2: Strip the search string of articles A, AND and THE.
Step 3: Tokenize the search input such that the input L is exploded to
        L1, L2, L3...:
        The input L = {w1w2w3…},
        Becomes L1 = {w1}, L2 = {w2}, L3 = {w3}…
Step 4: Build queries using a query builder.
Step 5: Use query to search through database for a result set.
Step 6: Use a ranking engine to arrange the result set according to the
result with the highest     frequency of a token.
Step 7: Produce the formatted form of the result.

Below is a structured English implementation of the search algorithm
SELECT keyword String
        REMOVE 'a' , 'and' , 'the' FROM string
        BREAK the keyword string into tokens
        FOR each token set
            SEARCH data storage
            IF token frequency > other token frequencies
                THEN print result
                ELSE
                Use earliest arrival Time.
END
```

Fig. 4. algorithm for the search process.

The algorithm for the content based recommender technique is depicted by Figure 5.

Both techniques are used in order to make user's result as useful and relevant as possible.

```
Start

        Step 1: Receive the tokenized string set.
        Step 2: Build queries using a query builder.
        Step 3: Search through data store using the
        user's preference as search key.
        Step 4: Use user preference to produce result.

End
```

Fig. 5. Algorithm for Content Based Recommender

Part of the search techniques applied is the krsort, which actually does the ranking.The techniques make the search process faster and the recommendation more reliable.

4 SYSTEM TESTING AND RESULT DISCUSSION

This section focuseson the testing and evaluation of the Academic Project Recommender System.

4.1 System Testing of APRS

The functionalities of APRS modules were validated to ensure its acceptability. Search and Recommending Module of APRS was tested. Search keys are entered in the search box and all project topics related to the search are returned as shown in figure 6. When the recommender button is clicked, it recommends a topic for the user based on the preference, that is, the search input entered on the search page.Also, figure 7is the snapshot showing the result of request for a recommendation – A recommended topic.

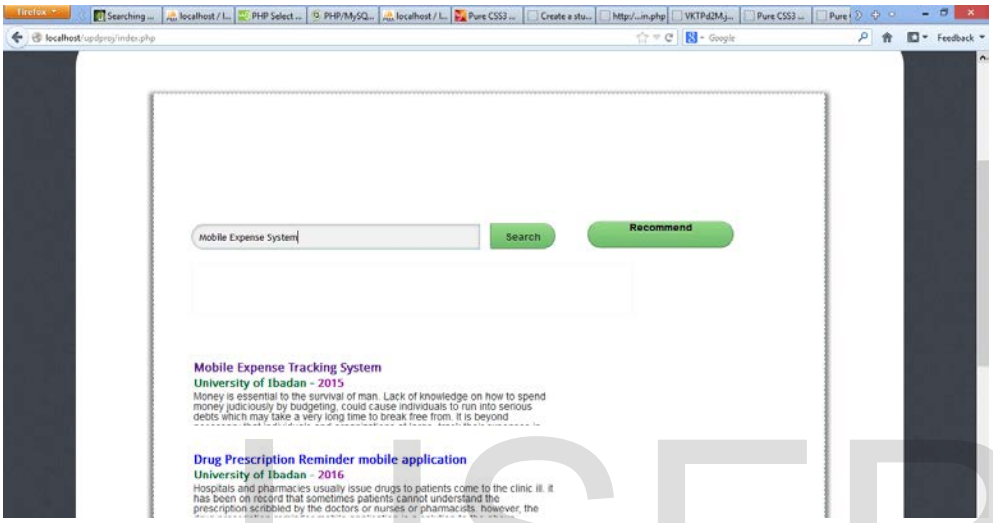

Fig. 6. Result of search query

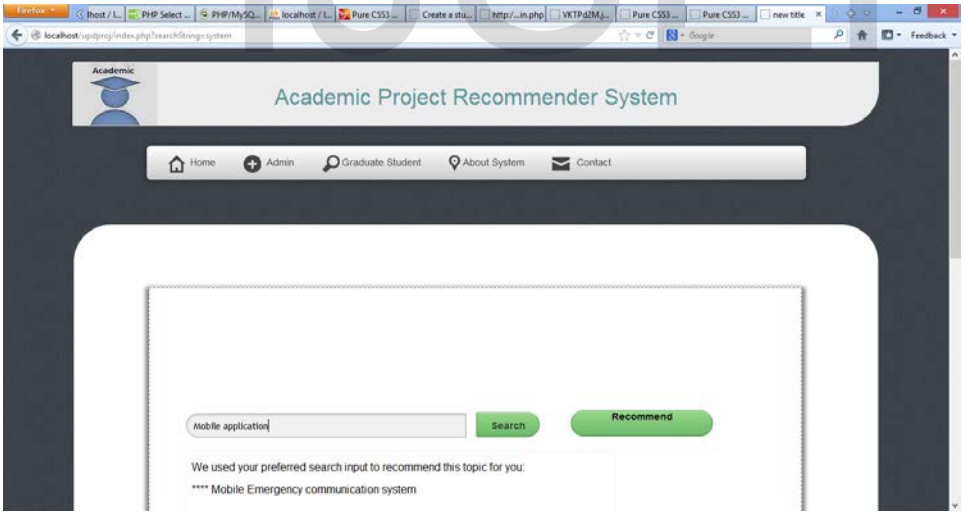

Fig. 7. The recommended topic

4.2 Evaluation of APRS and Existing Systems

This section reports the evaluation of APRS based on qualitative comparison with related works as highlighted in Table 1.

TABLE 1.
OMPARATIVE ANALYSIS OF APRS VS RELATED WORKS

	Academic Project Recommender System	Library Recommender System Based On Using the	Matchbox Recommender System	Research Based on Android Recommender Application	Amazon.com Recommendation
Technique	Content Based Recommender	Content Based Filtering	Bayesian method	Collaborative Based Filtering	item-item Collaborative Based Filtering
Algorithm	Map Reduce Variant	K-Means Algorithm	Learning Algorithm	k-means algorithm	item-item Collaborative Based Filtering
Aim	To develop a system that would help final year students to find viable Computer Science projects.	To develop a system that can reduce the workload and time of analyzing book borrowing transactions with respect to books borrowed, books that are always on demand or most preferred books.	To learn about people's preferences from observing how they rate items such as movies, content, or other products.	Research on collaborative based recommender systems and k-means algorithm with respect to a particular Android application.	To use recommendation algorithms to personalize the online store for each customer.
Limitation		only users' positive evaluations are available hence use of Jaccard's similarity coefficient		The system can get better if a newer filtering and recommending technology is adopted and applied.	Item-space partitioning restricts recommendations to a specific product or subject area.

5 CONCLUSION

The essence of the Academic Project Recommender System for final year students cannot be overemphasized. The system goes a long way in reducing stressful and almost futile project-topic-search processes, and to provide an avenue for research continuity. It recommends project topics for students to build upon.Further work is anextension to interoperate with search engines such that recommended topics are linked directly as search queries for further information retrieval.

ACKNOWLEDGMENT

The authors wish to thank Jessica Nwayanwu for her contribution to make this study possible.

REFERENCES

[1] J. Dewey, "Democracy and Education: An introduction to the Philosophy of Education." 1916. MacMillan Company USA.

[2] A.B. Fafunwa "History of Education in Nigeria", London, George Allen & Unwind Ltd.. Federal Republic of

Nigeria, 1981.National policy on Education. Revised. Lagos Nigeria, NERDC Press 1984.

[3] R.A. Aderinoye, "Expanding Access to Adult Basic Education in Nigeria: The intervention of Open Distance Learning. Adult Education in Nigeria" Vol. 14. 2007.

[4] Final Year Projects. Retrieved from Faculty of Engineering and architecture_ EECE:https://www.aub.edu.lb/FEA/ECE/STUDENTS/Pages/academics_fyp.aspx. 2010.

[5] Boolean Retrieval, "Cambridge University Press" 2009.

[6] J. B., Nicholas,,& W.C. Bruce, "Information filtering and information retrieval": Two sides of the same coin? 2002. Commumications of the ACM, 2

[7] P. Warden "Map Reduce". Retrieved from Pete Warden: http://petewarden.com/2010/01/20/mapreduce-for-idiots/2010.

[8] R., Francesco, L., Rokach, & B. Shapira, 2011. Recommender Systems Handbook.

[9] M. Uluyagmur, "Content Based Movie Recommendation Using Different Feature Sets". Proceedings of the World Congress on Engineering and Computer Science. 2012 Vol I.

[10] P., Lops, M., Gemmis, &, G. Semeraro "Content Based Recommender Systems". In P. Lops, Recommender Systems (pp. 76-77). 2012.

[11] M., Millan, M., Trujillo, & E. Ortiz, "Collaborative Recommender System". 2007. IDEAL, 633-672.

[12] Y. Chen, "Solving the Sparsity Problem in Recommender Systems". Journal of Computers, 1899-1900. 2011.

[13] M., Ekstrand, & J. Riedl, "Collaborative Filtering Recommender Systems". Retrieved from http://files.grouplens.org/papers/FnT%20CF%20Recsys%20Survey.pdf 2012.

[14] J. J. López, "The Importance of the Final Year Project as a way of Completing an Education in Engineering". International Conference on Engineering Education. Valencia. 2003

[15] P. Cosmos, "A Library Recommender System Based On Content Based Filtering Using the K-Means Algorithm". Nairobi. 2012

[16] D. Stern, "Matchbox: large scale online bayesian recommendations". NewYork. 2009

[17] B. F. Oladejo& I.A. Olaolorun, "Research Repository Management System (RRMS): A Research Repository Information System for Quality Undergraduate Research". International Conference on Science, Technology, Education, Arts, Management and Social Sciences, iSTEAMS.. Nigeria 2014.