A Suggested Algorithm of Recommender System to Recommend crawled-Web Open Educational Resources to Course Management System

Jamil A. Itmazi  
Director of e-Learning center  
Palestine Ahliya University (PAU)  
Bethlehem, Palestine  
j.itmazi@gmail.com

Haytham W. Hijazi  
Director of the computer center  
Palestine Ahliya University (PAU)  
Bethlehem, Palestine  
haitham@paluniv.edu.ps

Abstract — The majority of educational institutes and training centers are using some kinds of e-Learning via online platform i.e. Course Management System (CMS), Learning Contents Management System (LCMS). These platforms are somehow fixed to the e-contents of the tutor and teacher, while there are huge Open Educational Resources (OERs) available in the Web and ready for using and sharing. This paper proposes a Recommender System (RS) to recommend automatically OERs to a CMS after crawling them from Web to solve the students "Information Overload" problem arising from searching Web resources. This paper provides background of CMS - LCMS and RS as well as some examples. In addition, it discusses the suitability of main RS approaches to recommend digital resources from Web to support students’ needs. Finally, it presents a new proposal of RS algorithm which could automatically recommend suitable digital learning resources to a student in his active course.

Keywords— Recommender System, Open Educational Resources, Course Management System, Learning Management System

I. INTRODUCTION

Using web to deliver educational contents is the latest trend in training and education development industry [26]. However, the majority of the current online learning content management systems (LCMSs) are somehow fixed to the e-contents of the tutor and teacher. Moreover, the majority contents of the current LCMSs cannot be customized upon the learner’s interests.

A. Open Educational Resources (OERs)

We agree with whom calling Open Educational Resources (OERs) as a learning revolution since we see many big universities adopt OERs. Moreover, many other institutes started OERs Initiatives in the time we see some of OERs and Massive Open Online Courses (MOOCs) gain millions of students. Nowadays there are huge data related to OER ready for using, sharing, customizing. The main question here is how can we benefit from these huge OERs?

Many educational institutes and training centers are using an eLearning platform (LCMs) to automate the administration of their training events and educational contents. In the same time, the students suffer from the "Information Overload" problem, when they find in the Web thousands of results that are not suitable and not related to their LCMSs courses (eCourse).

Consequently, to solve this problem, the institute needs RS (Recommender System) to present interesting educational resources which relates to student's eCourse as well as fit their preferences from the internet via a topical web crawler (i.e., focused spider) that visits the web sites of OERs, crawl its contents, and index them based on the keywords. The retrieved resources could be listed automatically in a particular eCourse upon suitable priorities and ranking upon their importance to the tutor and learner. Figure No. 1 represents our RS system. Open Educational Resources (OERs).
Our web crawler will worth nothing if there are no open resources to crawl. Fortunately there are many educational websites allow the accessing and retrieving of their educational resource especially the OERs websites.

OERs began in 2001 when Massachusetts Institute of Technology (MIT) first announced to making all of its course materials freely available [16]. Its initiative called MIT OpenCourseWare (OCW) under the Website (http://ocw.mit.edu) to publish all of its undergraduate and graduate-level courses materials online, partly free and openly available to anyone, anywhere.

As of March 2014, over 2200 courses were available online [14].

The initiative has inspired a number of other institutions to make their course materials available as OER. Nowadays there are many institutions offer their course materials available online. We can find 40 of them in [6]. More and more could be found in [1]. As an example, the following projects and resources have been selected to illustrate the richness and diversity of the current initiatives in open educational and related resources and practices:

- M.I.T. Open Courseware (OCW), http://ocw.mit.edu
- Center for Open and Sustainable Learning (COSL) / OpenEd conferences, http://cosl.usu.edu
- Connexions (online platform for managing and sharing of open course modules), http://cnx.org
- Creative Commons, http://creativecommons.org
- Arabic Open programming school, barmaje.com

The term OER has been used to refer to learning materials such as:

- Learning objects (quizzes, animations, interactive maps, timelines, etc.)
- Audio lectures
- Audio-video lectures
- Images
- Sounds and music
- Entire course content
- Collections of journal articles and institutional repositories
- Textbooks

B. Course Management System (CMS)

It called also Learning Content Management System (LCMS) is an eLearning platform which is considered as an important part of eLearning solutions [9]. Moreover, there are some concepts similar to LMS (with a small difference), e.g. LMS (Learning Management System) and Portal Learning.

Generally, CMS is software that automates the administration of training events; it manages the log-in of registered users, manage course catalogs, track learner activities and results, as well as provide reports to management.

The market of CMS, LCMS and LMS is increasing fast, and there are hundreds of them in the market, for example [13] list 599 of them; some of CMS, LCMS and LMS are commercial Software, while others are free Open-Source LMSs. The following list shows some LMSs:

- Open-Source LMS: e.g. MOODLE <http://moodle.org> and ILIAS <www.ilias.de>.

C. Focused crawler

The e-course content has increasingly become a focus for the learning process. The educational contents’ update, however, has become a growing challenge to produce high quality and fresh educational material. Therefore, in this model we implement an effective web crawler to crawl the websites of OERs, e.g. http://ocw.mit.edu, download the
recent course-related educational resources, and integrate these updates with the existing course materials.

The web crawler [27, 4, 22, 28] is a web application that crawl websites by taking a list of seed URLs as an input, determine the IP address of the host name, download the corresponding resources and extract the links to continue the process. For example, the courseware Watchdog [23, 20] which is part of the Personalized Distrusted Learning Repositories (PADLR) framework has a focused or a topical web crawler to retrieve learning material from the WWW to be a part the learning materials.

In order to restrict the crawling to a material-relevant process, the crawling is done topically or focused.

II. RECOMMENDATIONS SYSTEMS

Have been widely implemented and accepted in many Internet sectors [5]. We are familiar with recommendations of products (e.g. books, music and movies) and of services (e.g. restaurants, hotels, Web sites); likewise recommendation is not arising from the digital era, but an existing social behavior in daily life. In everyday life, we rely on recommendations from others. [10].

Generally, the internet reaches the Billion Terabytes of data and the Web is still growing faster; as a result, the users suffer from the “Information Overload” problem, when searching the Internet [21]. Fortunately, the aim of RSs in Web applications is to present interesting information that fits the users’ tastes and preferences with little effort.

In contrast, sometimes RSs are used to hide special information, and specifically, the aim of RSs in eLearning applications is to present interesting information that fits the users’ tastes and preferences with little effort.

In the following we introduce the suitability of using RS approaches in the CMS.

A. Current Usage of RS

RSs have been widely used in many Internet activities. It is worth mentioning some examples of the current actual uses of RS:

- eCommerce: RSs are used “to suggest products to their customers and provide consumers with information to help them decide which products to purchase” [19]. eCommerce leaders like Amazon.com and Netflix have made recommender systems a salient part of their websites [11].
- Web pages: RS is used to solve the “overload problem” in the Internet, when using search engines (e.g. Google, Bing, yahoo) which produce thousands of pages to one researched item; most of them have worthless relation to the researched item or of no interest to the user. Example of search engines which used RS: Mi Yahoo! http://my.yahoo.com and Alexa.com.
- Censorship systems: RSs used to protect children from accessing undesirable material on the internet. e.g. cyberpatrol.com, as well as Prevent citizens from exploring some Web sites; which some governments already did.
- Other sectors: Examples:
  1) News: e.g. <www.lemonde.fr>,
  2) Encyclopedia: e.g. <http://wikipedia.org>,
  3) Software: e.g. <http://download.cnet.com>,
  4) Stores: e.g. <www.drugstore.com>,
  5) Tourist information: e.g. <www.viamichelin.com>,

B. RS and eLearning

eLearning is able to apply RS, which may be used to recommend the most appropriate content to students. In this paper, the focus will be at the use of RS in CMS. Some researchers mentioned the abilities of using RS in eLearning systems in general and CMS in particular. [3, 2, 12, 8, 18].

In the following we introduce the suitability of using RS approaches in the CMS.

III. THE SUITABILITY OF RS APPROACHES

RSs consist of approaches; every approach has its techniques. However, there are many systems that use Hybrid Recommender System, which combines two or more recommendation techniques to gain better performance.

Here, we are going to preview the suitability of the main RS approaches to recommend digital resources from Web to a CMS:

A. Content-Based System

In this type, the resources are selected by having correlation between the content of the resources and the user's preferences. In the context of this research this system could be used within RSs as a primary approach to find the digital resources from Websites, by detecting similarities between the current eCourse attributes (name, keywords, abstract …etc.) and the OERs attributes.

B. Collaborative Filtering Systems

It recommends items or resources to a target user, based on similar users’ preferences, and on the opinions of other users with similar tastes. It employs statistical techniques to find a set of users known as neighbors to the target user, examples: Amazon.com and ebay.com.
This system has some methods to calculate the likeliness from the rating matrix, the suitable one to our RS is Memory-Based Algorithm (also known as k-Nearest Neighbor Method), and because it is suitable to environments where the user preferences have to be updated rapidly.

C. Demographic-Based Filtering

It uses “prior knowledge on demographic information about the users and their opinions for the recommended items as basis for recommendations” [15]. It aims to categorize the user based on personal explicit attributes and make recommendations based on demographic group that a user belongs to, such as (income, age, learning level, or geographical region), or a combination of these clusters/groups. E.g. the Free e-mail suppliers put advertisements based on the user demographic information, like Yahoo as well as Google search engine.

The Demographic-Based Filtering could be used in the process of recommending digital resources as a complementary approach.

D. Rule-Based Filtering

It is filtering information according to set of rules expressing the information filtering policy [24]. These rules may be part of the user or the system profile contents and it may refer to various attributes of the data items. In general, this system is used widely with:

- Censorship: It is useful in the protection domain e.g. the protection of kids from accessing some materials, e.g. Cyberpatrol.com and Cybersitter.com [7].
- Spam Filtering: It is useful to be used against the Spam e-mails, e.g. Spam Assassin <http://spamassassin.apache.org> and MailEssentials <www.gfi.com>.

This system could be used within RSs to filter the recommendations list of digital resources upon some rules of the system and the student.

E. Hybrid Recommender System

“It combines two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one”. [17]. Examples of systems [21, 25].

IV. A GENERAL RS PROPOSAL

The suitable RS approach to recommend OERs from Web to CMS will not be a pure one, but it will be a Hybrid, which mixed some of the previous approaches.

The following general RS structure could be suggested (see Figure 2).

We list some consideration of this proposal structure:

- Content-Based System is used as a primary approach because it can give comprehensive, related and sufficient recommendations by using the resources attributes in the recommendation process.
- Collaborative Filtering Systems is not used as a primary approach because this approach becomes useful only after a "critical mass" of opinions, which means less numbers of recommendations or null recommendations.

This system could be used within RSs to filter the recommendations list of digital resources upon some rules of the system and the student.

D. Rule-Based Filtering

It is filtering information according to set of rules expressing the information filtering policy [24]. These rules may be part of the user or the system profile contents and it may refer to various attributes of the data items. In general, this system is used widely with:

- Censorship: It is useful in the protection domain e.g. the protection of kids from accessing some materials, e.g. Cyberpatrol.com and Cybersitter.com [7].
- Spam Filtering: It is useful to be used against the Spam e-mails, e.g. Spam Assassin <http://spamassassin.apache.org> and MailEssentials <www.gfi.com>.

This system could be used within RSs to filter the recommendations list of digital resources upon some rules of the system and the student.

E. Hybrid Recommender System

“It combines two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one”. [17]. Examples of systems [21, 25].

IV. A GENERAL RS PROPOSAL

The suitable RS approach to recommend OERs from Web to CMS will not be a pure one, but it will be a Hybrid, which mixed some of the previous approaches.

The following general RS structure could be suggested (see Figure 2).

We list some consideration of this proposal structure:

- Content-Based System is used as a primary approach because it can give comprehensive, related and sufficient recommendations by using the resources attributes in the recommendation process.
- Collaborative Filtering Systems is not used as a primary approach because this approach becomes useful only after a "critical mass" of opinions, which means less numbers of recommendations or null recommendations.

This system could be used within RSs to filter the recommendations list of digital resources upon some rules of the system and the student.

D. Rule-Based Filtering

It is filtering information according to set of rules expressing the information filtering policy [24]. These rules may be part of the user or the system profile contents and it may refer to various attributes of the data items. In general, this system is used widely with:

- Censorship: It is useful in the protection domain e.g. the protection of kids from accessing some materials, e.g. Cyberpatrol.com and Cybersitter.com [7].
- Spam Filtering: It is useful to be used against the Spam e-mails, e.g. Spam Assassin <http://spamassassin.apache.org> and MailEssentials <www.gfi.com>.

This system could be used within RSs to filter the recommendations list of digital resources upon some rules of the system and the student.

E. Hybrid Recommender System

“It combines two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one”. [17]. Examples of systems [21, 25].

IV. A GENERAL RS PROPOSAL

The suitable RS approach to recommend OERs from Web to CMS will not be a pure one, but it will be a Hybrid, which mixed some of the previous approaches.

The following general RS structure could be suggested (see Figure 2).

We list some consideration of this proposal structure:
The eCourse items include: name, keywords, abstract... etc. Empirically, as shown in Figure 3, the tutor configures the crawler for his own course -- CSS is the example here. First, the tutor is supposed to enter the seed URL/URLs for his course, and devised the suitable keywords for the course content. Second, he determines the successive crawling frequency. And finally, the type of fetched content is selected.

Fig. 3. Crawler content panel

Consequently, as shown in Figure 4, the crawler visits the source file of the specified URL (e.g., search.mit.edu), the recognizes the CSS as a keyword, and fetch the URLs under the mata tag "<a href="..." as a potential content resource.

Fig. 4. Source file of search.mit.edu

B. The Stage of Teacher Recommendations

The teacher has the ability to upload digital resources of his own, colleagues, OERs ... associated to an eCourse into the Repository of Digital Resources. These Resources are recommended ones and mainly the best ones.

After the teacher uploads any resource, the system is deleting any duplicate. Also, the system gives the teacher resources high priorities.

C. The Stage of Filtering

When the students inter his eCourse, this stage will be activating. This Stage consists of:

- Collaborative Filtering: this approach will organize the priorities of the recommendations. The general mechanism based on defining subgroups (every subgroup known as the nearest neighbors) whose preferences are similar to the active user, so the nearest neighbors of the active student are those students who share the same institute (department, school). Then this stage calculates the average of the subgroups rating to order the recommendations upon the high rates.

- The Rating Matrix: The target CMS need to have a way to capture the rating by explicit, implicit methods or mixture of them. These students' rates of the digital resources saved in the CMS database as a table of two dimension matrix; where the row represents all the rates of one student on all digital resources while the column represents all the rates of all students on one digital resource (see table I).

Table I. rating matrix

<table>
<thead>
<tr>
<th>Student</th>
<th>Dr1</th>
<th>Dr2</th>
<th>...</th>
<th>Drm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std1</td>
<td></td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Std2</td>
<td>2</td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>...</td>
<td>3</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Stdn</td>
<td>3</td>
<td></td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

- The general steps of Collaborative Filtering are:
  a) Receiving the list of the recommended digital resources.
  b) Finding the neighbors of the active student.
  c) Finding the average rates of the neighbors for every recommend digital resource.
  d) Organizing the recommendations upon the highest average; firstly, organizing the set of the “teacher recommendations” which already have the higher priorities then organizing the other recommendations set which came from Web.
  e) Finally, the “recommended digital resources” are passed to the next steps.
- Demographic-Based Filtering: Theoretically, the role of this approach is to filter the incoming recommendations upon the students’ demographic (and personal) data that related to education issues. For example, the following demographic-personal data could be related to the education issues: preferred language, student specialization, study level year, faculty, and department. The language filtration as an example means that the active student needs all the recommended digital resources in his preferred language, so any language of digital resources in the recommendations list defer from his preferred language will be deleted.

- The general steps of Demographic-based Filtering are:
  a) Receiving the list of the recommended digital resources.
  b) Reading the related demographic and personal data of the active student profile.
  c) Matching the related fields of each digital resource from the list with the fields of the active student profile, so if the matching process is not positive; the digital resource will be deleted from the list.
  d) Finally, the “recommended digital resources” are passed to the next steps.

- Rule-Based Filtering: It will filter the incoming recommended digital resources upon a set of rules, which could be found in the student profile and in the system profile. The system administrator put some rules in the system profile, while the student can put his own rules in his profile.

We suggest that the following types of rules that could be used in the student profile and the system profile to filter the listed digital resources (see figure 5):

1) Link: the system will filter out any digital resource whose link found in the rules profiles.

2) Phrase or word: the system will filter out any digital resource which his name, keywords or abstract match any “phrase” or “word” found in the rules profiles.

3) Date: the system will not show any digital resource does not fit the date criteria.

4) Size: the system will not show any digital resource does not fit the size criteria.

5) Type: the system will not show any digital resource does not fit the type criteria.

- The general steps of Rule-Based Filtering are:
  a) Receiving the list of the recommended digital resources.
  b) Reading the following fields of the system rules:
    1) Field which contains link of digital resource.
    2) Field which contains keywords.
    3) Fields of maximum and minimum dates.
    4) Field which contains the allowed size.
    5) Field which contains the forbidden types.
c) The system deletes from the recommendations list every digital resource that matches any link or keywords as well as any digital resource whose dates are out of the minimum-maximum dates. It also deletes any digital resource, whose size is larger than the allowed size and whose type matches the forbidden types.

d) Reading the same fields of rules from the student profile and repeating the filtration process.

e) Finally, the “recommended digital resources” are prepared to be presented in a suitable way on the windows of active student eCourse.

V. DISPLAYING RESULT VIA CMS/LCMS

Displaying results to students could be done within many methods. As a case, it could be displayed via eLearning platform (LMS, CMS and LCMS). Respect to our paper, the appropriate platform is the open source one which allow adding new part within the course page to display the results when the tutor/student accessing his course. Fortunately there are many Open-Source LCMSs, e.g.

- MOODLE <http://moodle.org>
- ILIAS <www.ilias.de/ios/index-e.html>
- Claroline <www.claroline.net>

As an example, we could use MOODLE in our case.

The procedure could be the following:

a) When the user (student or tutor) logging into his eCourse, Moodle retrieves the existing resources in the repository which related to the current eCourse.

b) The RS activate the stage of Filtering and finishing with a list of recommended digital resources.

c) The list is showing in a “block” in the eCourse page.

Figure 4 shows this example block within Moodle course page.

VI. CONCLUSION

RSs have been widely used in many Internet activities, mainly to overcome the information overload problem, which the user faced while searching any item and getting thousands of unrelated results. This research tries to solve the overload problem when students searches Web about suitable and related digital resources to their current eCourse as well as preferable resources to their needs and taste.

This research summarizes some essential information about crawler system, CMS/LCMS and RS. In addition, it reviews the suitability of RS approaches to recommend digital resources from Web to CMS/LCMS. Furthermore, the paper studies and presents a new RS algorithm to recommend suitable digital OERs from Web to students while entering an eCourse in CMS/LCMS. These proposed algorithm is considered as a Hybrid Recommender System which consist of some RS approaches; Content-Based System, Collaborative Filtering, Rule-Based Filtering and Demographic-Based Filtering.

REFERENCES


