Search Engine Repository for the Personalization of Learning for University Students: A Case Study in Guadalajara, Mexico

J. López-Cisneros[1(✉)] and A. L. Franzoni-Velaquez[2]

[1] Blvd. Marcelino García Barragán, Guadalajara 44430, México
juan.lopez@academicos.udg.mx
[2] Progreso Tizapan, Río Hondo 01080, México
analidia@itam.mx

Abstract. University students participating in a course face the challenge of choosing educational resources when the suggested ones are not suitable. Educational trends encourage the use and development of technologies to improve learning processes. New technological proposals suggest the development of adaptive and/or recommendation (personalization) capabilities according to the user's specific needs. The proposed methodology focuses on the development of a web repository for exploring the recommendation of educational resources and evaluating its implementation. Two algorithms were implemented based on the publication of educational resource references. The first relates to the reference search engine, while the second focuses on analyzing user interactions with the references using an unsupervised model. The study concludes with an examination of students' perceptions as the web repository progresses. The web prototype was used for activities designed throughout an academic cycle with 115 students from the institution. Students perceived progress in retrieving references on the web platform as they advanced through the academic cycle, providing feedback on performance, content, and visualization improvements.

Keywords: Educational Resources · Personalization · Web Platform · Deep Learning

1 Introduction

When designing the academic content for a school subject, there is the opportunity to choose the educational resources and tools that will facilitate and/or complement a student's understanding of a subject. The increase in Internet services provides users access to resources in quantity, diversity, and timelessness. A common behavior from students, when the resources proposed in the course are not useful enough for their learning, is to retrieve them from other places according to their specific needs.

There is a growing interest in two areas that are focused on the incorporation and exploration of meaningful data capabilities in education: Educational Data Mining

P. C. López-López et al. (Eds.): ICOMTA 2023, AHSSEH 11, pp. 238–247, 2024.
https://doi.org/10.2991/978-94-6463-254-5_24

(EDM) and Learning Analytics (LA) and their respective communities [1]. Furthermore, that adaptive/personalized learning has become a fundamental learning paradigm in the community of educational technology research.

Personalized education has been defined as providing a new educational model to meet the individual desires of students and the development of society [2]. Also, that it has been a focus in the education area and developing new methodologies for personalized education is very urgent.

Assessing the relevance of educational resources in the training of students is a challenge for an academic. First, they need to have a reference store for educational materials; second, they have to analyze students' interactions with said educational resources; finally, they need to consider a strategy for resource personalization in the face of current educational requirements.

Although there are technological platforms that support these educational processes, certain difficulties exist in their adoption. Some causes are the possible restricted access to the platforms, the retrieval of open data from users' information transactions generated for their studies, and, finally, the educational vision implemented by their creators.

A situation that arises in Web information systems is the problem of delivering to each type of user the content that is truly of interest to them [3]. It also considers the strengthening of personalization techniques supported by the formation of user groups and the adaptation processes associated with them, which have been a case of study mainly in the collaborative learning area.

The purpose of this work is to support educational processes and to obtain accurate information to carry out the corresponding data analysis. For this reason, this type of tool has been created to obtain 'a priori' data to carry out studies on the behavior of users with the educational resources available on the Web and, where possible, generate personalization.

2 Related Work

The use of technology allows for a flexible pace; personalized instruction; immediate interventions; anywhere, anytime during learning; and offers students control of their own learning [4]. Examples of how technology is currently used to implement this are: learning management systems (LMS), educational resource repositories, recommendation systems, and search engines.

Moodle is one of the most used LMS in educational institutions. Personalization within a Moodle course can be implemented using access restrictions at the activity or resource level [5]. The resources available to students in the course are added manually by the course manager. When any of them are accessed, interaction data is automatically stored for possible further analysis by a process external to the platform's operation.

One instance of an educational resource repository is Merlot [6], a platform created by California State University. The repository identifies, describes, evaluates, and links with online learning objects. Access is free and it is designed for higher education activities. Due to the amount of information contained in the repository and in order to improve their search processes, it was determined to discover more effective approaches to automatically classify educational resources into relevant discipline categories [7].

Applied in various fields of work, either on commercial websites like Amazon or entertainment products like Netflix, recommendation systems are present in the education field. They are built with components (learning models) that allow adaptation in time or preferences evolution. For example, one recommendation proposal is based on the perception of the utility of the educational resource through collaborative filtering voting and the K-nearest neighbors (KNN) algorithm [8], while in another work, the concept of semantic association is adopted to suggest educational resources as a result of exploring Linked Open Data environments [9].

The Internet is recognized as a reliable source of information and services in different domains. It implies that as web retrieval methods became more efficient and effective, the Internet gained the trust of users significantly and quickly [10].

Search engines are a tool for retrieving information on the web. Given a focus on the recovery of educational and research resources, they can be found as a component in various platforms such as Google Scholar, Web of Science (WoS), Scopus, Massive Open Online Courses (MOOC), and Teacher Tube. As an example, in Google's new policies, when a person searches for information on their search engine, the search does not happen in isolation. Their previous activities are tracked on the web and used to train personalization algorithms [11].

The application of recommendation (personalization) in the field of education requires taking into account a wide range of variables, such as the level of knowledge, skills, and learning styles of the students. Given the rapid evolution of these systems, it is important to be attentive to the trends in the techniques used for their development [12].

3 Method Design

The methodology used in the present paper is based on an investigation for educational resources personalization, which involves the following phases: 1) Definition of the architecture and logic of the platform (see Subsection A); 2) recovery of educational resources (see Subsection B); 3) machine learning technique (see Subsection C); 4) evaluation of the web platform (see Subsection D); 5) the case study (see Subsection E). The description of each phase is given in the following subsections.

3.1 Definition of the Architecture and Logic of the Platform

For the network and information technologies to truly serve the education and optimize the process, it is necessary to have rich teaching information resources to support it [13].

The web platform proposed for the personalization of educational resources is defined by two aspects: the first is in regards to its general architecture and the second is with its logic.

The construction of an educational resources database is the central focus of educational informatization [13]. Figure 1 shows the simplified schema of the repository's database and its main entities.

Each type of educational resource to be registered on the platform determines a series of attributes or metadata to be stored. For this, an "attribute-value" structure is

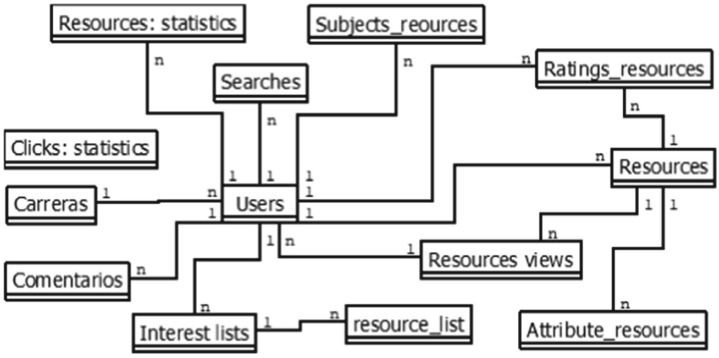

Fig. 1. Entity-relationship scheme of the proposed database.

considered in the database with the intention of making the increase in content types more flexible.

Figure 2 shows the platform's architecture [14], where development components can be observed, such as the web platform interface, the database, the result search algorithm, and the generation of the results view of them.

This proposal has been viewed as a reference repository that supports students and teachers in their academic activities. As a reflection, in a similar study, it is stated that Learning Objects (LO) and Open Educational Resources (OER) emerged from a real need to easily share knowledge with accessibility and availability [15]. The platform intends to facilitate references to educational resources.

As a first step, the user makes the required registration and authenticates through the web interface. For the second step, the interaction operations that users can perform, such as searching for resources, qualifying them, and also labeling in which subjects educational resources can be used, have been established. Finally, the generation of interest lists has been added for the grouping of educational resources related to a particular learning strategy.

Figure 3 shows a simplified user diagram, explained in the preceding paragraph, that contains the interaction made for the activities allowed on the web platform.

One of the support strategies in the registration of educational resources onto the web platform is the use of "content managers". Users who intervene as validators and

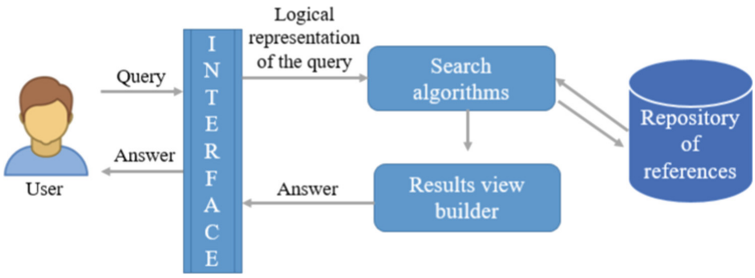

Fig. 2. Web platform's structure.

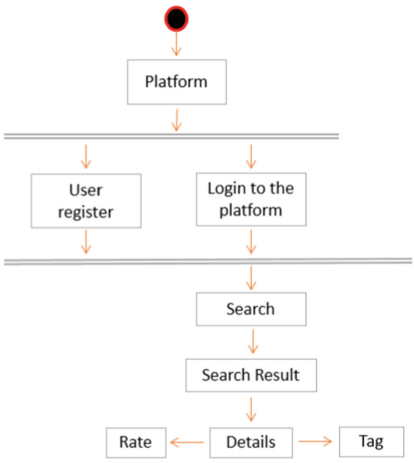

Fig. 3. User activity on the platform.

curators. This strengthens the idea of sharing the resources used by discipline experts which may generate a collaborative learning experience.

3.2 Educational Resource Recovery

The adoption of alternative ways of analyzing and interpreting knowledge by information retrieval systems is required, where the user-system interaction is guided by intrinsic elements that increase the chances of being efficient in information retrieval [16].

A starting approach to educational resources personalization is beginning with the recovery of the database elements using operations in a structured query language (SQL). The results are then sorted by a different algorithm which will establish the ordering in which resources will be returned to the user. This sorting algorithm uses automated scoring on the resource from different variables with the intent of personalizing them.

To get this scoring the following variables and weighting (arbitrary values) are considered:

- If any word from the search string is in the "title" a weight of "1" is assigned.
- The number of times that the word "matched" in the different attributes (description, link, author, etc.). That "match" is multiplied by "0.5".
- The average from the user's scoring in stars given to that resource.
- The resource type that is set to preferred (text, video) determined with a machine learning technique.

A second approach is the recovery of interest lists created by the content managers and students. An independent search component was implemented through SQL consultations. The strategies to be used in generating a personalized list for the students are studied.

User							...		
	2	5	2	5	4	5	...	4	1
	5	4	3	5	3	4	...	5	2
	5	2	4	4	3	5	...	5	1
	3	1	4	5	1	4	...	4	3

Fig. 4. Example of the dataset for applying clustering.

3.3 Machine Learning Technique

Users and resources are stored in the database. The use of the platform through a web interface allows for a series of interactions that are also stored. This is done in order to apply collaborative filtering.

In collaborative filtering recommendation systems, approaches based on item and user are used. That is, it is a way to filter or evaluate an item based on the thoughts and opinions of other people. [17].

When retrieving the recorded interactions of the resources, a collaborative filtering of all the records is performed to generate another dataset (Fig. 5). In the proposed application, a machine learning technique is applied, specifically unsupervised learning. With a clustering algorithm, users are characterized to determine the type of resources they prefer. The initial algorithm used for this version is k-means.

Starting from 'a priori' knowledge base, to which unsupervised learning algorithms are applied, it is important to consider the negative effects of issues such as a "cold start" and the "increase of new users" when there is insufficient information available on the platform.

3.4 Web Platform Evaluation

In order to evaluate the web platform, it was proposed to perform three work stages over a scholar cycle, which would be:

The first stage is a usability study performed at the beginning of the scholar cycle. It consisted of the online application of 4 questionnaires to the students, which consisted of a consent form, a form to identify the users, the use of the platform being evaluated, and a satisfaction survey through a "Computer System Usability Questionnaire (CSQU)".

Second stage, activities (school assignments) were designed for the students. These activities entailed the search for resources and the generation of interactions on the web platform. After completing each activity, feedback on the experience was submitted.

The third stage, a questionnaire was applied to evaluate the course at the end of the scholar cycle, in which for the ends of this study, an evaluation on the pertinence of the activities (school assignments) made with the support of the platform was included. Said questionnaire included at the end an open feedback section about the use of the platform over the course.

Fig. 5. Web interface view of an educational resource.

3.5 The Case Study

The case study was developed with two objectives. First, to evaluate the platform functionality through the three stages specified in the preceding section. Second, to identify the improvement opportunities thinking ahead on the development of a second version.

Figure 5 shows a view of the interface from the first prototype of the web platform used in the "Applied Programming" course imparted in the institution.

For the first phase, its implementation was carried out in three out of the eight active student groups of that subject from August 10th to December 15th, 2021. The number of students that could potentially interact with the platform was 115, between students coursing the Electrical Mechanics Engineering and Industrial Engineering majors.

For the other two stages, the study group was reduced to one of the three groups that initially interacted with the platform. The number of participants in this second stage was 35 students of the previously indicated majors.

Due to the restrictions in place from covid-19 pandemic, implementation was made remotely with the support of a virtual learning environment, online forms, and e-mail use.

4 Results

In order to evaluate the proposed web platform, the results are presented for each stage.

In the first stage of 115 students between three groups, only 80 answered in full the 4 web questionnaires. From these, in user identification, it was encountered that 79% of the surveyed are in the age bracket between 18 to 20 years old. 75% prefer accessing

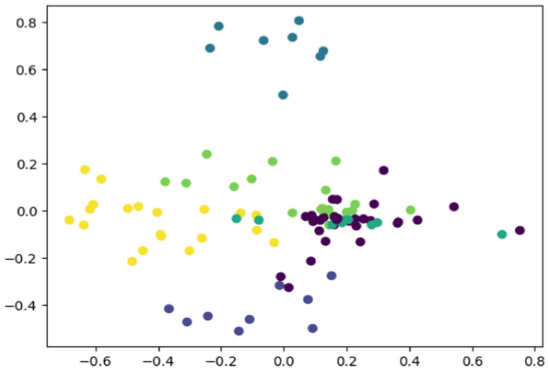

Fig. 6. Application of the K-means algorithm to the user-educational resources dataset.

an educational platform from a desktop computer. 57% of the users make use of some educational platform for more than 30 h a week.

In regard to the evaluation and the satisfaction survey, the obtained results were that 70% of the users qualified the score as simple and easy to use, 63% expressed being able to perform the tasks and 35% manifested themselves in a positive manner on the use of the interface.

For the second phase, out of the total of 35 students, only 28 students consistently participated in the completion of the 10 proposed activities. This allowed for the observation of functionality errors and feedback on the perception of result delivery by the search algorithm.

For the third phase, the study concludes with the administration of the course evaluation questionnaire at the end of the academic term through a Google form. From the received feedback, some positive comments were, for example, about how useful it is when searching for specific things related to the career; it acts as a filter to separate deficient resources from good resources; on the other hand, there were mentions of the lack of content in certain topics and the need for a more visually appealing platform.

Regarding the analysis under an unsupervised model, 751 resources were stored in the repository during the proposed period. By constructing the feature matrix, it was found that 94 users interacted with 87 stored resources. When applying the K-means algorithm, as shown in Fig. 6, the possibility of clusters generation can be observed.

It should be mentioned that those columns (features) in the feature matrix that were only indicated once by a user were removed.

5 Future Work and Conclusion

The proposed web platform helps with recovering educational resources and references that support university students. From the consultation of these resources, the interactions were stored to build data sets and apply the generation of the clusters as unsupervised learning technique.

Over the course of a scholar cycle, its functionality was studied from the first working version. 70% of the students agreed that the platform was simple and easy to use,

but pointed that in the future simplicity should be kept, content increased and visual experience improved.

In the application of the machine learning algorithm, once the number of clusters is determined, a qualitative analysis of each group is required to determine their preferences. The need for more data for a comprehensive exploration is emphasized.

In the next scholar cycle, after gathering the needed authorizations, a new iteration of the three stages mentioned in this paper will be applied. In this one, more professors will be invited to use the platform on their educational practice. The number of educational resources will be increased in many dominions, and interest lists for sharing will be generated.

All of this with the ends of having more interactions in the platform, generating new data sets, exploring new algorithms, and make improvements in personalization.

References

1. Shemshack, A., & Spector, J. M.: A systematic literature review of personalized learning terms. *Smart Learning Environments 7*(1), 1-20 (2020). https://doi.org/10.1186/s40561-020-00140-9
2. Xiao, M., & Yi, H.: Building an efficient artificial intelligence model for personalized training in colleges and universities. *Computer Applications in Engineering Education 29*(2), 350-358 (2021). https://doi.org/10.1002/cae.22235
3. Cobaleda, L. & Duitama, F.: Personalization of content in adaptive educational hypermedia systems: a review, *Revista Facultad de Ingeniería Universidad de Antioquia* 50, 217-226 (2009).
4. White, H. E.: Personalized learning (2014). [Online]. Available: https://www.learningpersonalized.com/wp-content/uploads/2015/02/E-Personalized-Learning-Revised-FINAL.pdf
5. Papanikolaou, K., & Boubouka, M.: Personalised learning design in Moodle. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT, IEEE)* 57–61 (2020). https://doi.org/10.1109/ICALT49669.2020.00024
6. Schell, G. P., & Burns, M.: Merlot: a repository of e-learning objects for higher education. *E-Service* 1(2), 53-64 (2002). https://doi.org/10.2979/esj.2002.1.2.53
7. Gunarathne, W. K. T. M., Shih, T. K., Chootong, C., Sommool, W., & Ochirbat, A.: An Automated Learning Content Classification Model for Open Education Repositories: Case of MERLOT II. *Journal of Internet Technology 21*(5), 1277-1288 (2020).
8. Garnica, C. C., Archundia-Sierra, E., & Beltrán, B.: Prototype of a Recommendation System of Educative Resources for Students with Visual and Hearing Disability. *Res. Comput. Sci. 149*(4), 81-91 (2020).
9. Duran, D. F., Chanchí, G. E. & Arciniegas, J. L.: Educational resource recommendation system based on learning goals and reasoning in linked open data, *Revista Ibérica de Sistemas y Tecnologías de Información* E32, 1-13 (2020).
10. De Medio, C., Limongelli, C., Marani, A., & Taibi, D.: Retrieval of educational resources from the web: A comparison between Google and online educational repositories. In *Advances in Web-Based Learning–ICWL 2019: 18th International Conference, Magdeburg, Germany, September 23–25, 2019, Proceedings* (pp. 28–38). Cham: Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-35758-0_3
11. Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z.: Measuring political personalization of Google news search. In *The World Wide Web Conference,* 2957–2963 (2019). https://doi.org/10.1145/3308558.3313682

12. Urdaneta-Ponte, M. C., Mendez-Zorrilla, A., & Oleagordia-Ruiz, I. Recommendation systems for education: systematic review. *Electronics*, 10(14), 1611. (2021). https://doi.org/10.3390/electronics10141611
13. Wang, X., & uhn Jong, K.: Design and Analysis of Educational Resources Recommendation Service System Based on 5G Era. In *2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI, IEEE)* 114–117. (2020). https://doi.org/10.1109/ICAACI50733.2020.00029.
14. Tolosa, G. H. & Bordignon, F. R.: Introduction to Information Retrieval, 12 (2008). (Unpublished) [Book] [Online]. Available: http://eprints.rclis.org/12243/
15. Colome, D.: Learning objects and open educational resources in higher education, Edutec. *Revista Electrónica de Tecnología Educativa* 69, 89–101. (2019). https://doi.org/10.21556/edutec.2019.69.1221
16. Cuba Rodríguez, Y., & Olivera Batista, D.: Metadata, search and information retrieval from the Information Science. *E-Ciencias de la Información 8*(2), 146–158. (2018). https://doi.org/10.15517/eci.v8i2.30085
17. Zheng, Z. IoT Applications for Recommended Methods of Physical Education Online Course Resources Based on Collaborative Filtering Technology. Wireless Communications & Mobile Computing. 1–9 (2022). https://doi.org/10.1155/2022/2928559