Recommendation system with graph-oriented databases for repository of open educational resources

A. Morales¹, J. Gonzalez¹, K. Alquerque¹, J. Reyes¹ and S. Sanchez¹

¹Antonio José de Sucre University Corporation (Colombia).

jaider_reyes @ corposucre.edu.co

Abstract. The advancement of telecommunications has stimulated individualized study and collaborative learning, allowing the creation of pedagogical innovations by the teaching and administrative staff of educational institutions. Among these innovations are the Open Educational Resources (OER), which since its inception have contributed to teaching practices and contributed to the free learning of students. For this reason, the objective of the research is to design a repository of open educational resources integrating a recommendation system with graph-oriented databases. The study is descriptive and technological development. The results are the schematization of the system architecture by means of a block diagram and flow diagram.

1. Introduction

Globalization contributes to the arrival of an interactive society, but it can also cause secondary effects, especially with regard to access to information and knowledge in cyberspace. Therefore, international institutions and the competent national authorities of each country must develop strategies to mitigate this problem. UNESCO proposes that public access to information and its free circulation should be promoted without geographic, economic or social discrimination, through appropriate public policies that guarantee its reliability. [1]

The advancement of telecommunications has stimulated individualized study and collaborative learning, allowing the creation of pedagogical innovations by the teaching and administrative staff of educational institutions. Among these innovations are the Open Educational Resources (OER), which since its inception have contributed to teaching practices and contributed to the free learning of students [2].

The development of a repository for educational resources is undoubtedly an excellent tool for the teaching-learning process of educational institutions, but it should be noted that the growth of information is exponential and, therefore, better approximate and match the information that it is consulted of interest, it becomes more complex and tedious. Today there is the need to select between several alternatives without having a fixed knowledge of each of them, in these situations, the final decision may depend on the recommendations made by other people, for this reason currently companies such as Spotify, YouTube, Amazon guide users through recommendations.

These systems have been called "Recommendation Systems". These systems are in charge of providing users with personalized and differentiated information on certain topics, products and / or services that may be of interest. That is, they are responsible for guiding the user through recommendations in the search for those topics, services or products that may be more attractive, modifying the browsing or search process. The use of these systems is becoming increasingly popular.
due to their usefulness in evaluating and filtering the large amount of information available on the Web. This is undoubtedly a great advantage for users, who will find what they need in a faster, more comfortable and easier way within electronic repositories on the Internet and also allow them to discover new topics.

Therefore, the main objective of this research is to design a repository that integrates a recommendation system with the use of graph-oriented databases, which facilitate the recognition of patterns in the relationships that join the nodes. The database will relate user searches by topics of interest in order to make personalized recommendations.

2. Recommendation Systems

With the arrival of Big Data, the massive amount of information available on the web creates the need to make decisions when making inquiries and selecting products, content or topics of interest, however frequently due to not having the necessary information or sufficient experience staff on the subject, use is made of recommendations from other people (family or friends) or by the media such as Radio or TV [3].

At the beginning of the 90's, developers and computer science researchers, faced with this situation, developed recommendation algorithms, which in the last decade have taken center stage in many of the popular computer applications such as, for example, online stores, movie repositories, video repository, music repositories, academic databases, among others. The purpose of these recommendation systems is to satisfy the search needs of users in a personalized way [4].

Recommendation systems have been added to e-commerce websites (such as Amazon.com) and leisure applications such as Netflix, Spotify, YouTube, Facebook, Instagram, Telegram, Snapchat, among others that have made these systems a prominent part of the provision of services [5].

Recommendation systems are tools whose main objective is to help users in making decisions to select the item that best suits their preferences, tastes and interests. [6], the recommendation systems can also be defined as common solutions in the personalization of content, they identify elements that are of interest to the user using a recommendation component that compares the characteristics defined and stored in the user profiles, to predict the score of the Internet user towards items you have not yet considered [7].

These systems have characteristics in which the system presents elements that different users with similar tastes like, the system forms a user profile according to the interaction that the user has had with the elements and recommends content based on the knowledge acquired. [3], so that the system already has the ability to select the items that the user might be interested in and the user is recommended to recommend topics, products or services that resemble other products that they have liked. [8].

2.1 Recommendation Technique

It can be said that these systems select the useful elements for the user among a large set of existing search elements, that is why these systems must anticipate the possible decisions that the user is going to select. For this, the system uses different methods and algorithms to calculate the recommendations, among the different algorithms there are the content-based, the collaborative filtering and the hybrid hybrids. [9].

2.1.1. Content-based algorithm. In this type of algorithm, the context is any information that can be used to characterize the situation of the entities (person, place or object) that are considered relevant for the interaction between a user and an application, including the user and the application itself. Themselves. Context is typically the location, identity, and state of people, groups, and physical or digital objects. [10]

2.1.2 Collaborative Filtering Algorithm. This type of algorithm is a method to make predictions about the interests of a user based on information about the preferences of other users. Collaborative filtering
is the process of filtering information or models, which uses techniques that involve collaboration between multiple agents, data sources, among others.[11]

2.1.3 Hybrid algorithm: The hybrid filtering technique combines various recommender techniques to obtain a better optimization of the system to avoid some limitations and difficulties of the pure recommender systems. The idea behind hybrid techniques is that a combination of algorithms will provide more accurate and effective recommendations than a single algorithm, since the disadvantages of one algorithm can be improved by another algorithm. The use of multiple recommendation techniques can suppress the weaknesses of an individual technique in a combined model [12].

3. Graph-Oriented Database
Graph-oriented data stores information through the structure of graphs, where the information is grouped into nodes and these in turn are related by memory pointers, which allows these databases to apply graph theory for navigation; These are very useful for storing information in models with many relationships such as social networks and connections, where each node is assigned a grade that indicates the number of connections it has[13].

Since a high demand for information is required when storing it and when searching for information in the application when making inquiries, it leads to investigating new technological tools that facilitate data management. Taking into account that the data provided by the user must be studied and interpreted in order to be useful, therefore, if we could represent the data and the relationships between objects as a single data set, knowledge patterns could be generated that describe the set of data contained in these elements together. For this purpose, the databases are presented in graphs that are flexible and powerful enough to represent these elements.[3].

3.1 Neo4j
Neo4j is a native graphics database, built from the ground up to take advantage of not only data but also data relationships. Neo4j connects data as it is stored, enabling queries never before imagined, at speeds never thought possible [14].

It is a robust (fully ACID) highly scalable graph-oriented database. Neo4j is used in mission critical applications by thousands of leading companies and governments around the world.

Below in the figure, 1 presents a diagram of what would become "NEO4J", as an implementation in a graph-oriented database [3].

![Figure 1. Function representation of neo4j.](image-url)
Neo4j Main Features

- Intuitive: use a graphical data model for data representation.
- Reliable: complete ACID transactions.
- Durable and fast: native storage engines.
- Highly scalable: thousands of nodes, relationships and properties.
- High availability: distribution across multiple machines.
- Expressive: powerful query language.
- Embedded
- Simple: accessible with a REST interface or an object-oriented API (JAVA).

4. Methodology

Research is of a technological development type, according to MinCiencias this type of study occurs in the application of the results of the research, or of any other type of scientific knowledge, for the design or development of new processes, systems or provision of services, as well as the substantial technological improvement of pre-existing materials, products, processes or systems. This activity will include the materialization of the research results in a plan, scheme or design, as well as the creation of non-marketable prototypes [fifteen]. The purpose of this research is to design an OER repository for the Antonio José de Sucre University Corporation to offer educational content to the student community and teachers of the department of Sucre located in Colombia.

To carry out the research, two phases were developed using the Cascade software development methodology (see figure 2):

![Figure 2. Cascade Methodology.](image-url)

In the requirements phase, a bibliographic review was carried out and with it a documentary analysis of the information provided by the expert entities in charge of the development and directors of virtual learning environments, and of academic documents results of the investigation of related topics recommendation systems in any field, be it tourism, commerce, entertainment or education, in order to identify the functional and non-functional requirements of the system proposed in the research.

For the design phase, the requirements identified in the previous phase were taken into account and from then on, through the use of flow diagrams and block diagrams, the system architecture was
designed, relating the communication between the different technologies that integrated the system. system operation.

5. Results
Due to the continuous development of digital content, it is necessary for the repositories databases of these contents to have scalability characteristics as they grow, in addition to having the capacity to facilitate the storage and retrieval of information. However, the massive amount of information leads to users not always finding the desired information, and if they do, the task becomes complex and tedious. That is why recommendation systems play an important role in solving and managing this information within these repositories.

Within the research, the design of the recommendation system for a repository of educational resources for the Antonio José de Sucre University Corporation, was made initially a bibliographic review was made organized in the bibliographic matrix and analyzed by means of an analytical content matrix, later The requirements survey was carried out through semi-structured interviews applied to experts in the area, these requirements are listed in table 1.

| Functional | Not functional |
|------------|----------------|
| The system should allow searching for topics of interest to the student. | The system interfaces should be easy or intuitive. |
| The system must allow modification of the search. | |
| The system should provide assistance when conducting a search. | |
| The system must list the users according to the topics of interest. | |
| The system must identify which topics are of interest to the user or not. | |
| The system must make recommendations according to the user's profile. | |
| The user must view the results of the recommendation according to interests. | |
| The user must view independent results in case of no history. | |

As a proposal after the analysis of the proposed requirements, the following software architecture was designed, which expresses each of the processes that are necessary for the proper functioning of the recommendation system. The tests will be carried out with a NEO4J graph-oriented database engine.
This system will give recommendations of topics to each user individually according to the interest of each one of them, thus optimizing searches, permanence on the site and most importantly, it is the optimization in the performance of the same system, since This affects the user's perception of their learning process.

As can be seen in the flow diagram figure (see figure 3), the system starts with a search, which returns a list of results that concatenates with search history, in addition the system provides recommendations of the most viewed or consulted topics by their peers. The system identifies a new user who provides independent results according to the search topic, results that will be saved in their profile for a future recommendation.
Figure 4. Scheme of more generalized recommendation system above.

As can be seen in figure 4 "block diagram", it is a more generalized representation of the recommendation system previously analyzed. The process begins in the search for educational resources, then goes to the first block, which receives the search data, to later continue to the next block, where the data of queries made in the user's history are extracted, followed by this a comparison of the data from the new search with the history, checking similarities between the two. If you do not have any history as a user, it will present you with independent results according to the topic consulted, the system makes a comparison of the resources seen and not seen to record it as history and topics of interest for future comparisons and then generates new recommendations.

6. Conclusions

It is concluded that the implementation of recommendation systems has had a boom in recent years, because they allow the handling of information to be more precise for end users, that is, these systems prevent the user from being frustrated and lost on the web due to the massive amount of information that exists in cyberspace.

It is also concluded that recommendation systems have been a great pivot for electronic commerce and marketing, to such an extent that they have proven to be a tool of great value for organizations that have implemented them within the Internet and their consumers, with the aim of that users continue to visit the site and increase the probability that they carry out some commercial activity. Due to these positive results in this sector, it is necessary to carry out research on the implementation of these systems in other areas to evaluate their impact.

Regarding the results of the research, it was possible to analyze and design an architecture capable of generating recommendations for educational resources from a historical set of data, information that will be in constant growth due to the relationships that will be formed according to the interests of each user, thus creating patterns that help the system to better predict in each recommendation process; It is
worth mentioning that, to make a prediction with a minimum degree of error, it is inevitable to analyze a large amount of data.

References

[1] UNESCO, "Access to Information," [Online]. Available: https://es.unesco.org/themes/acceso-informacion. [Last access: August 2020].

[2] E. Nila, C. Juárez, J. Sánchez and A. Soberanes, "Open educational resource design with an expert system for database teaching-learning," Research in Computing Science 111, pp. 23-32, 2016.

[3] WAV Vargas, «Context-sensitive recommendation algorithm for reusable educational elements with Graph-Oriented storage,» Revista Tecnológica ESPOL - RTE., Vol. 30, No. 1, p. 101, 2017.

[4] VHMDAZG Rodrigo Vences Nava, «Performance optimization of a text document recommendation system based on server configuration,» COMPUTACIÓN E INFORMÁTICA, n° 2, 2016.

[5] SGMB Ajay Kaushik, "A Movie Recommendation System using Neural Networks," International Journal of Advance Research, Ideas and Innovations in Technology, p. 425, 2018.

[6] CAR Morales1, «SVD algorithm applied to recommendation systems in commerce,» Tecnología, Investigación y Academia, vol. 6, no. 1, p. 19, 2017.

[7] DVDYML José Carlos Hernández Nieto, "NAPI: NEWS PERSONALIZATION SYSTEM," VI INTERNATIONAL SYMPOSIUM ON COMMUNITY, ENTERTAINMENT AND SOCIAL MEDIA "., 2018.

[8] LDM-CLCR-LV and. S.-AMA Ibarra-Báez, «Book Recommendation System,» Applied Computational Technology, vol. 1, No. 1, 2019.

[9] P.-y. LRW and. Z.-f. Z. Pei-pei Wang, "A Recommendation Algorithm Based on Density Peak Clustering and Key Users," 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017), 2017.

[10] TGJABCAMTGBV Paola Gómez, «Recommendation Systems: an approach to filtering techniques,» Revista Ibérica de Sistemas e Tecnologias de Informação, p. 288, 2019.

[11] XLWDRZ FULIAN YIN, "TAG-BASED COLLABORATIVE FILTERING RECOMMENDATION ALGORITHM FOR TV PROGRAM," 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2016.

[12] W. Wenzhen, "Personalized Music Recommendation Algorithm Based On Hybrid Collaborative Filtering Technology," International Conference on Smart Grid and Electrical Automation (ICSGEA), p. 281, 2019.

[13] C. Pinilla, M. Bello and C. Peña, "Graph-oriented databases,» Tecnología, Investigación y Academia, vol. 5, no. 2, pp. 153-160, 2017.

[14] Neo4j, «Neo4j,» August 13, 2020. [Online]. Available: https://neo4j.com/. [Last access: 13 August 2020].

[15] MinScience, Typology of Projects Classified as Scientific, Technological and Innovation Characteristics, Bogotá: Government of Colombia, 2019.