Journal of Universal Computer Science, vol. 24, no. 11 (2018), 1515-1535
submitted: 26/2/18, accepted: 7/9/18, appeared: 28/11/18 © J.UCS

EduRP: an Educational Resources Platform based on Opinion Mining and Semantic Web

Maritza Bustos López
(Tecnológico Nacional de México/I. T. Orizaba, Av. Oriente 9 No. 852, Col. E. Zapata
94320 Orizaba, Veracruz, México
maritbustos@gmail.com)

Giner Alor-Hernández
(Tecnológico Nacional de México/I. T. Orizaba, Av. Oriente 9 No. 852, Col. E. Zapata
94320 Orizaba, Veracruz, México
galor@itorizaba.edu.mx)

José Luis Sánchez-Cervantes
(CONACYT-Tecnológico Nacional de México/I. T. Orizaba
Av. Oriente 9, No. 852, Col. E. Zapata, 94320, Orizaba, Veracruz, México
jlsanchez@conacyt.mx)

María del Pilar Salas-Zárate
(Tecnológico Nacional de México/I. T. Orizaba, Av. Oriente 9 No. 852, Col. E. Zapata
94320, Orizaba, Veracruz, México
pilarsalasz12@gmail.com)

Mario Andrés Paredes-Valverde
(Tecnológico Nacional de México/I. T. Orizaba, Av. Oriente 9 No. 852, Col. E. Zapata
94320 Orizaba, Veracruz, México
mapv1015@hotmail.com)

Abstract: Educational platforms have become important tools for e-learning; nonetheless, finding the appropriate educational resources to use often represents a tedious task for learners. Opinions in the educational domain are important information for decision making; they allow teachers to improve the teaching process and enable students to decide on the best educational resources. The large amount of data that is daily generated on the Web makes it difficult, however, to analyze opinions manually. Multiple opinion mining approaches are being proposed as a solution to this problem; this research work introduces EduRP, an education platform that integrates opinion mining techniques and ontology-based user profiling techniques. We specifically propose an opinion mining approach for Spanish text which consists of three main steps: 1) collect opinions from the EduRP platform, 2) process the opinions to normalize the text, and 3) obtain the polarity of the opinions using a machine learning approach. We also propose a profile customization approach that uses Semantic Web technologies, specifically ontologies, to integrate socio-demographic data from different social networks and from the platform itself. Finally, we assess the performance of our system under precision, recall, and F-measure metrics, obtaining average values of 81.85%, 81.80% and 81.54, respectively.

Keywords: Machine Learning, Natural language processing, Opinion mining, Sentiment Analysis
Categories: I.2.7, I.7, H.3.3, L.3.2

1 Introduction

Educational resources are useful tools to support student learning. According to [Anido et al. 2002] an educational resource is an entity that can be used or referred to through a learning process. Multimedia contents, textbooks, tutorials, workbooks and quizzes are examples of educational resources. Nowadays, due to the amount of content generated by users on the Web, finding appropriate resources for student learning is a tedious task. It is therefore necessary to apply techniques that allow students to filter the educational resources that have obtained the best assessment by other users and finally obtain personalized recommendations based on their user profiles. By reducing the time and effort that is required to find suitable educational resources, students can focus more on the learning itself. At the same time, the application of these techniques can help teachers to know how students have evaluated available educational resources and take a decision to improve them if required.

Opinion mining and ontologies are important tools to deal with the above-mentioned problem. According to Liu, opinion mining is a technology that allows analyzing people's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities, such as products, services, organizations, individuals, issues, events and topics, and their attributes [Liu 2015]. One of the main tasks of opinion mining is polarity identification, which is commonly carried out with the aim of classifying users' feedback on resources into positive, negative and neutral feedback. Recent trends in polarity identification research in education focus on two main approaches: semantic orientation identification and machine learning. The former approach involves the use of lexicons such as SentiWordNet [Baccianella, Esuli and Sebastiani 2010] to determine the polarity of the words contained in a text. As regards the machine learning approach, two data sets are required: one for training a classification algorithm and one for testing the resulting predictive model [Salas-Zárate María del Pilar, Paredes-Valverde Mario Andrés, Limon-Romero Jorge, Tlapa Diego and Baez Lopez Yolanda 2016]. Even though both methods have provided satisfactory results, multiple works have demonstrated that machine learning approaches are more effective. Ontologies represent one of the main building blocks of the Semantic Web; they are complex and formal vocabularies that semantically describe the concepts and relations used to represent a specific domain. Ontologies enable the sharing and reuse of static knowledge, thus reducing the effort needed to implement knowledge-based systems.

In this work, we propose EduRP, an educational resources platform that integrates opinion mining techniques and ontology-based user profiling techniques. Our contribution is therefore two-fold. On the one hand, we propose an approach for machine learning-based opinion mining on Spanish text, whose aim is to analyze comments of EduRP platform users. The Spanish language is the second most spoken language in the world and the third most used language on the Internet; nonetheless, the development and application of automatic polarity identification systems in non-English languages (including Spanish) are rarely explored by Human Language Technologies (HLT) or Natural Language Processing (NLP) researchers [Martínez Cámara 2016]. On the other hand, we propose a semantic profile development approach, which involves socio-demographic information retrieved from social networks and our platform itself. Socio-demographic information is crucial in the educational domain, because it allows teachers, instructors, and facilitators to determine the impact of socio-demographic factors on both student academic performance and attitudes toward e-learning experiences.

The structure of this paper is as follows. Section 2 presents a review of the literature on opinion mining and ontology-based user profiling in the educational domain. The overall design of our approach is described in Section 3, whereas Section 4 introduces a case study of the generation of opinion mining statistics for students and teachers. Section 5 describes the experiments and the evaluation results concerning the effectiveness of the proposed system. Finally, conclusions and remarks about future work are presented in Section 6.

2 Related work

2.1 Opinion mining

Opinion mining is an ongoing research field in computer science, and it is also a fast-growing research field with applications in multiple domains. In educational contexts, the benefits of opinion mining techniques are explored due to the importance of forums, blogs, social networks, review websites, and other Web 2.0 tools for both learners and teachers. Therefore, multiple opinion mining methods, algorithms, systems, and platforms have been proposed from both lexicon-based and machine learning based approaches. The levels at which these approaches deal with the polarity classification task range from the feature/aspect level to the document level, including word and sentence levels.

Authors [Kravvaris and Kermanidis 2017] proposed a semantic-based information retrieval technique for machine learning-based opinion mining, whose aim is to improve educational video retrieval. The technique was evaluated using a dataset of educational YouTube videos to filter the videos' comments judged as negatives that had a weak semantic connection with the videos' verbal content. To this end, the polarity of the comments was computed at the sentence level. On the other hand, [Rajput, Haider and Ghani 2016] presented a lexicon-based method for text analysis of end-of-course student evaluation feedback at a higher education institution. The goal of the method was to compute the polarity of free-text evaluation comments at the word and sentence levels.

Opinion mining techniques have also been proposed as algorithms in the context of recommending systems, adaptive e-learning systems, and student feedback systems. For instance, an e-leaning recommending system for teachers was proposed in [Tewari, Saroj and Barman 2015] with the aim of timely improving the quality of the e-materials from the learners' viewpoint. The system relies on a feature-based opinion mining approach that uses a lexical resource to identify parts of the subject topics that are difficult to understand by learners. From a different perspective, [Kechaou, Ammar and Alimi 2011] introduced a pioneering research on the applications of opinion mining in education and a machine learning-based opinion mining approach for reviews from e-learning blogs and forums. The approach seeks to improve the quality of e-learning systems from a development perspective, and to this end, it addresses polarity classification at the document level.

[Ortigosa, Martín and Carro 2014] proposed an opinion mining approach for Facebook posts whose objective is to highlight regular patterns and changes on user sentiments. The approach can be integrated into adaptive e-learning systems by defining learner models incorporating sentiment information, thus enabling the recommendation of motivational activities for groups and individuals. Unlike other approaches, this one is a lexicon-based machine learning hybrid approach. In [Dhanalakshmi, Bino and Saravanan 2016], the authors developed a machine learning-based opinion mining approach in the context of a student feedback system at a higher education institution. The goal was to automatize learning analytics tasks on the data from the evaluation surveys. Similarly, the work proposed to calculate the polarity of the survey responses at the sentence level. On the other hand, [Leong, Lee and Mak 2012] proposed a feature-based opinion mining approach for SMS messages from an SMS response system at a high school. The work sought to timely improve the quality of any of the aspects of the lectures at the school from the viewpoint of students. For that purpose, the system separately computes the polarity of all the concepts extracted from the messages.

As regards opinion mining systems and platforms, authors [S. Rani and Kumar 2017] proposed a system for temporal analysis of student feedback. The system classifies the polarity of messages from an SMS response system and comments from the Coursera e-learning platform. Then, it classifies the sentiment associated to each Coursera comment using a lexicon-based approach to highlight teacher performance and compute learner satisfaction/dissatisfaction scores for courses. Furthermore, [Zarra, Chiheb, Faizi and Afia 2016] proposed a cloud-based opinion mining platform for e-leaning whose objective was to benefit from the exchange of messages among students from different institutions when solving blockages and gaps in community courses. The platform relies on a lexicon-driven machine learning hybrid approach to classify the polarity of posts from discussion forums at community clouds, such as StackOverFlow.

2.2 Semantic profiling

Semantic profiling is a useful and recurrent technique in the development of (personalized) recommendation systems and (adaptive) e-learning systems. In the first case, semantic profiling is commonly used to represent knowledge about the user and about the object of decision (from a decision support systems perspective), or the document (from an information retrieval perspective). In the second case, semantic profiling is typically used to represent knowledge about the learner and the learning resources. The array of semantic profiles that can be built for users or learners in these contexts is vast, but demographic and social profiles, as well as domain-dependent preference profiles, are commonly built and exploited.

Semantic Web ontologies are formal explicit specifications of shared conceptualizations that can be seen as the natural enabler for the semantic profiling technique. In this context, the concept of "domain ontology" has become popular among knowledge-based systems. It refers to the use of an ontology to provide a common vocabulary for a domain and define the meanings of the terms and the relationships between them. Domain ontologies can be seen as a means to build semantic profiles for the items that are the object of the recommendation; however, it is a common practice in system design to also include the terms and relations involved in user characterization.

In [M. Rani, Nayak and Vyas 2015], the authors proposed a cloud-based adaptive and personalized e-learning system that focuses on capturing learning styles rather than on delivering learning resources. The system uses one ontology to represent knowledge about learners and another to represent knowledge about courses, topics, and learning materials (i.e. the domain ontology). Similarly, in [Sharma and Ahuja 2016], the authors discuss a knowledge-based collaborative filtering recommending system for e-learning content. The system integrates an ontology-based semantic similarity technique and a similarity metric for user-based top-n recommendation. One ontology that models computer science topics was used to represent the e-learning content in the system, whereas a second ontology was used to represent knowledge about the learner. Namely, the learner ontology was used to build hybrid user profiles containing social information about learner preferences.

A knowledge-based recommender system for e-learning objects was introduced in [Tarus, Niu and Yousif 2017]. The system integrates an ontology-based semantic similarity metric and a sequential pattern-mining algorithm. The authors propose one ontology to build learner profiles containing demographic data (i.e. age and genre), as well as other types of information, such as learning styles and knowledge levels. Then, a second ontology is used to represent knowledge about the e-learning objects, namely object types (e.g. exams and assignments) and object formats (e.g. video and text). Unlike other works, this system effectively allows building semantic profiles for the items that are the object of the recommendation. From a different perspective, [Heiyanthuduwage, Schwitter and Orgun 2016] presented an e-learning ontology that comprises e-learning concepts, learning object metadata, and terms and relations for characterizing learners. The ontology is intended to be easily integrated into e-learning systems and thus to facilitate the access to e-learning resources. In fact, the authors suggested including the concept of learner profile directly in a domain ontology.

Finally, in their work, Alimam, Seghiouer, & Elyusufi developed a software architecture of an e-learning and career guidance hybrid system for middle school students [Alimam, Seghiouer and Elyusufi 2014]. The system uses a single ontology to represent knowledge about learners and e-learning content. Furthermore, it includes concepts and relations for the characterization of learner profiles. In this case, however, learner profiles are characterized in accordance with the professional interest categories of Holland's model of careers and vocational choices.

Table 1 depicts the results of a comparative analysis that we have carried out between related works. It summarizes relevant properties of these pieces of research in terms of four main aspects: 1) language; 2) approach; 3) linguistic resource; and (4) data. As can be observed, machine-learning-based opinion mining approaches rely on supervised machine learning. In this context, examples of the most popular classifiers include Support Vector Machines, Naïve Bayes, C4.5 (decision tree), and k-Nearest Neighbors. Conversely, in lexicon-based opinion mining, available lexical resources range from academic (e.g. MPQA Subjectivity Lexicon) to commercial (e.g. LIWC analysis tool). Finally, notice that only one of the approaches discussed above deals with opinions in Spanish.

	Language	**Approach**	**Linguistic resource**	**Data**
[Kravvaris and Kermanidis 2017]	English	Supervised machine learning	Corpus 1000 opinions (not available)	Youtube
[Rajput et al. 2016]	English	Lexicon-based approach	MPQA Subjectivity Lexicon	Forums, social networks
[Kechaou et al. 2011]	English	Supervised machine learning	Corpus 2000 opinions (not available)	Blogs and forums
[Ortigosa et al. 2014]	Spanish	Combined approach	Linguistic Inquiry and Word Count (LIWC) Corpus 3000 opinions (not available)	Facebook
[Dhanalakshmi et al. 2016]	English	Supervised machine learning	Corpus 6433 opinions (not available)	Module Evaluation Survey (MES)
[S. Rani and Kumar 2017]	English	Lexicon-based approach	NRC Emotion Lexicon	Coursera course
[Zarra et al. 2016]	English	Supervised machine learning	Corpus 120 post (not available)	Forums

Table 1: Results of the comparative analysis.

In conclusion, opinion mining applications in the education domain promote the development of e-learning systems. Moreover, opinion mining approaches are an area of opportunity to better capitalize subjective information. As demonstrated in the end of this section, opinion mining approaches can take advantage of semantic profiles not only to represent knowledge, but also to reason about that knowledge. In this sense, the development of knowledge resources for opinion mining, which is a pending issue in opinion mining research, can be particularly valuable in the education domain.

3 Architecture

This section discusses the architecture and functionality of our platform. The architecture is divided into four main layers as depicted in Figure 1: the presentation layer, the data analysis layer, the processing layer, and the data collection layer. The presentation layer ensures communication between users and the system. This layer is composed of a set of graphical user interfaces (GUIs) developed with HTML5

(Hypertext Markup Language, version 5), CSS3 (Cascading Style Sheets Level 3), and JavaScript. On the other hand, the data analysis layer comprises two modules: opinion mining and semantic profiling, whereas the processing layer consists in the normalization of text. Finally, the data collection layer includes two modules: one to extract the comments, and the other to extract user socio-demographic information from social networks and our platform itself. A detailed description of each layer and its components is provided in the following section.

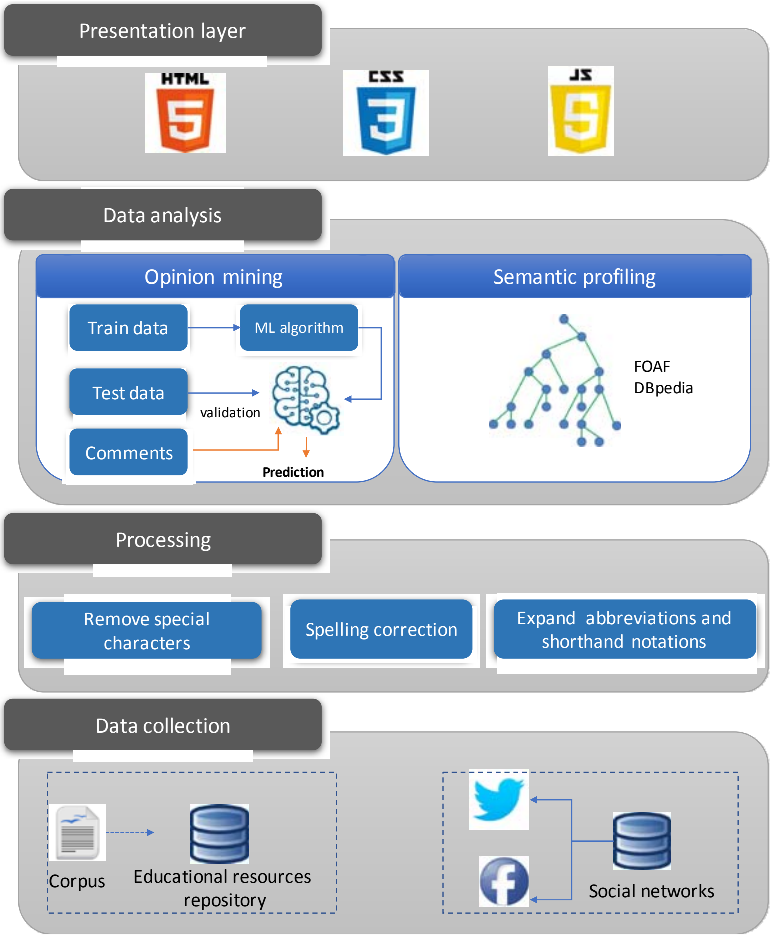

Figure 1: System architecture.

3.1 Data collection and processing

Data collection layer extracts information from both the platform and social networks. It is composed of two main modules:

1. **Data social networks**: This module obtains information from social networks. We use Facebook Graph API [Facebook 2018] and Twitter API [Twitter 2018] to respectively extract information from Facebook and Twitter. We specifically obtain data from user profiles, such as name, given name, family name, gender, age, religion, among others.

2. **Data platform**: This module extracts comments on the educational resources available on the platform. A corpus is compiled from these data to be used in the opinion mining module. In this work, two experts in education reviewed and classified each opinion into three different categories: positive, negative and neutral.

Since the language used on the Web is commonly informal, the texts must be normalized after being correctly processed by the other modules. The processing layer involves the use of natural language processing techniques to normalize the text. Three specific tasks are performed:

1) Remove special characters that do not provide relevant information for the opinion mining and semantic profiling module. In this step strings such as URLs are removed.

2) Expand abbreviations and shorthand notations by their expansions. Abbreviations and acronyms are commonly used on the Web. For instance, the word *también* (meaning *also* in English) is usually, but informally, abbreviated as *tb.* We relied on the NetLingo [Jansen 2014] dictionary in this step.

3) Spelling correction. We used the spell checker and morphological analyzer Hunspell [Németh 2005] to correct spelling errors, such as "profesr," "bueon," and "mathematics," which were replaced by "profesor" (*professor* in English), "bueno" (good), and "matemáticas" (mathematics), respectively.

The tools employed for these tasks have been successfully used for text normalization in other works [Salas-Zárate, Paredes-Valverde, Rodriguez-García, Valencia-García and Alor-Hernández 2017].

3.2 Opinion mining

This module obtains the polarity of each comment made on the educational resources. To this end, it adopts a machine learning approach (see Figure 2) in which two data sets are required: a training set and a testing set. The former is used to train the machine learning algorithm and build the predictive model, whereas the latter is used to evaluate the performance of the obtained model. This model allows classifying new comments into positive, negative, or neutral.

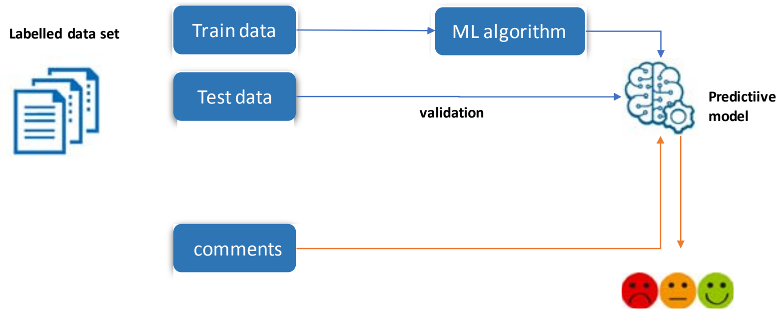

Figure 2: Machine learning approach

The opinion mining module was implemented by using the Natural Language API [Google 2018], which can analyze texts and classify writer attitudes into positive, negative, or neutral. The overall opinion (positive, negative, or neutral) in a text is determined by a numerical score and a magnitude value. The score represents the general emotion of the text, while its magnitude value represents how much emotional content the text includes. In this sense, neutral polarity can indicate low-emotion in the text or both, positive and negative opinions. Magnitude values are useful in disambiguation. For instance, truly neutral texts will obtain a low magnitude value, while mixed texts will obtain higher magnitude values. Table 2 depicts two examples which obtained a score of 0, i.e. a neutral value. The first one represents, however, a mixed text because it obtained a magnitude value of 1.2. This magnitude value indicates mixed emotions with both high positive values and high negative values that cancel each out. The second one represents a truly neutral text because it obtained a magnitude value of 0, which indicates a low-emotion document.

1	The educational resource is well explained. However, the presentation is very bad	Score: 0 Magnitude: 1.2	Mixed text
2	The resource contains examples of linear algebra	Score: 0 Magnitude: 0	Neutral text

Table 2: Examples of neutral texts.

3.3 Semantic profiling

Social networks are important sources of information. Users rely on them to share their personal information, activities, and thoughts, among others. In e-learning systems, social network data are valuable resources to build personalized user profiles, know user preferences, and identify the impact of socio-demographic factors on e-learning attitudes. The semantic profiling module of our system builds semantic user profiles with socio-demographic information. As depicted in Figure 3, the module utilizes semantic technology to integrate data from the platform itself and from both Facebook® and Twitter®. In this sense, the architecture uses an ontology to model learner features, which was designed by taking both the friend of a friend

(FOAF) [Brickley and Miller 2000] and the DBpedia [Auer et al. 2007] ontologies, since the two contain vocabulary for describing people and activities.

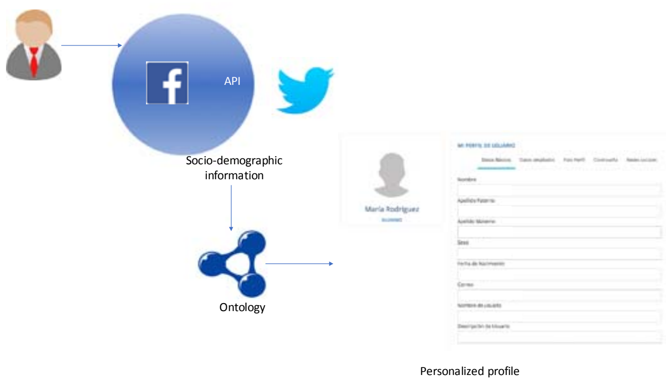

Figure 3: Personalized profile

As regards FOAF, the *name*, *givenname*, *familyname*, *title*, *age*, and *gender* properties were considered. On the other hand, the *religion* and *profession* properties from DBpedia were selected. Also, the ontology is described using the W3C RDF Schema and the Web Ontology Language (OWL) 2. Figure 4 illustrates an excerpt of the ontology, whereas the following paragraphs describe its concepts.

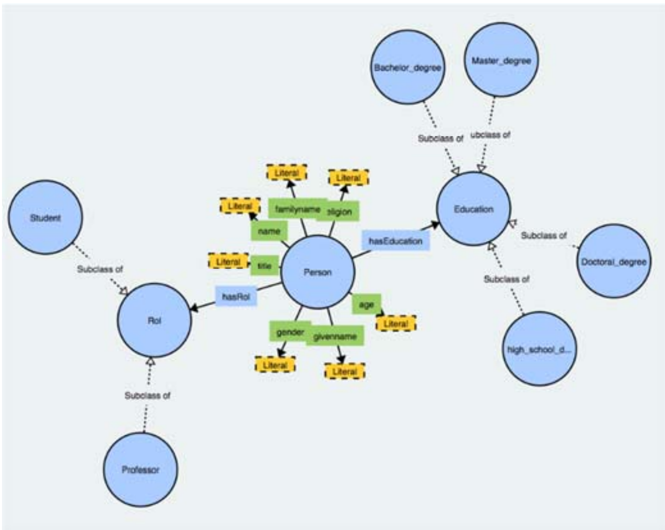

Figure 4: Ontology excerpt

- Person: It represents people and contains properties such as given name, family name, age, and gender, among others.
- Role: It represents the role of a person: i.e. a student or professor.
- Education: It represents a person's school degree (high school, bachelor master, PhD).

4 Case study: Generating opinion mining statistics for teachers and students

The main goal of our system is to stand out as a supporting tool in teaching-learning processes through opinion mining techniques. In order to exemplify the operation of the platform, this section introduces two case studies: one for students and one for teachers.

4.1 Generating opinion mining statistics for students

Learners usually consult educational resources on the Web, yet they might spend a considerable amount of time searching for the appropriate information. Our system relies on opinion mining techniques for learners to know, and thus have access to, the most useful resources (according to student reviews). Similarly, the system displays the overall opinion of a specific resource.

1. Once students have logged in, the system shows a graphical interface (see Figure 5) to search for educational resources with respect to three criteria: keywords, category, and type. Category refers to the area of knowledge, including mathematics, Spanish, biology, geography, history, physics, chemistry, arts, sports, English, and ethics. Meanwhile, type refers to the format in which the resources are provided (i.e. text, image, audio, video, word document, PDF, power point, and flash). The search results can be visualized in a grid view, and the interface displays information such as resource title, author names, description, average rating, visualization counter, and comment counter.

2. By clicking on the "Details" button, students can consult additional information on the resource. Then, the system displays the interface illustrated in Figure 6, which is divided into four sections. 1) The resource details section displays resource title, description, topic, author, and publication date. Then, 2) at the bottom of the resource display frame, the interface displays the resource's average rating, a five-star rating option, the view counter icon, the comment counter icon, and the Favorite icon. Additionally, (3) the comment section allows students to write their review on the consulted resource. Finally, (4) the opinion mining section graphically displays the percentage of positive, negative, and neutral reviews.
3. When selecting the "Opinion mining" option, the system displays a graph with the percentage of positive, negative, and neutral reviews for a resource. Also, students can visualize the top ten educational resources (see Figure 7), according to their search.

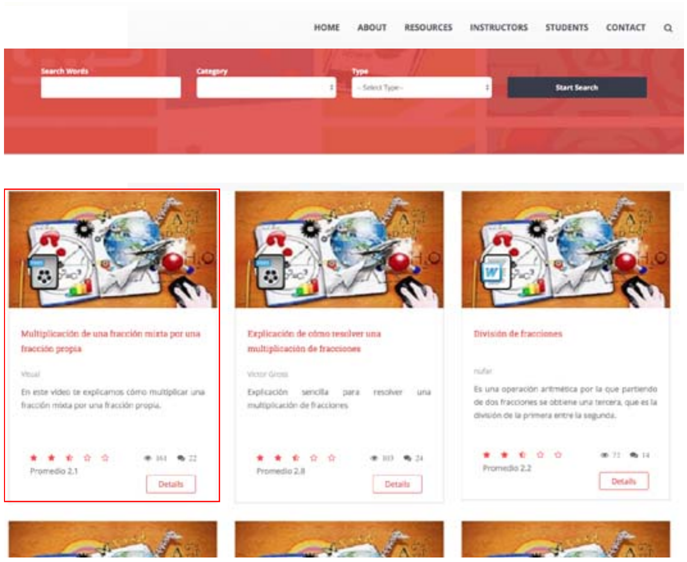

Figure 5: Educational resources in the system.

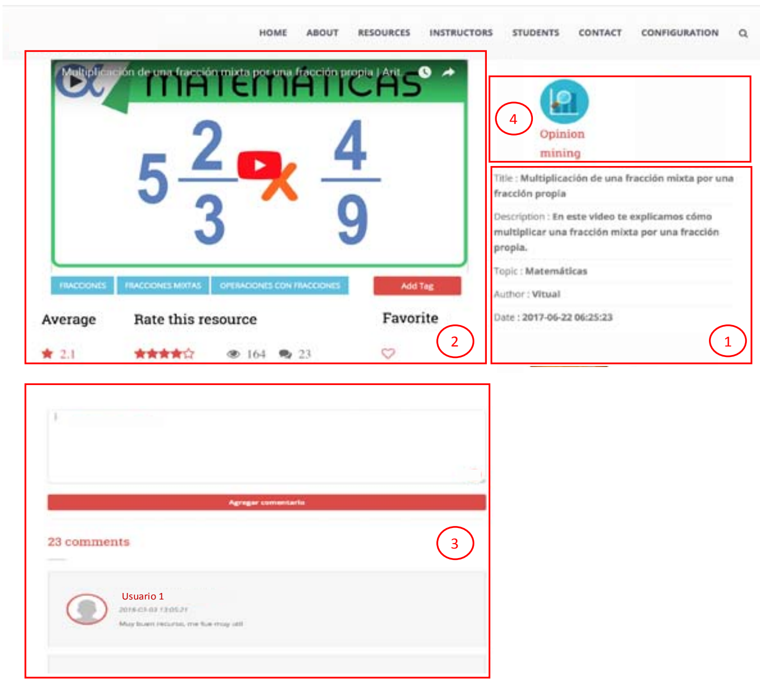

Figure 6: Resource details.

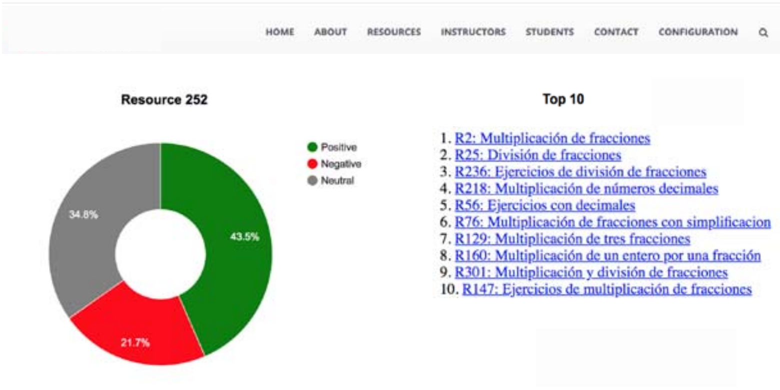

Figure 7: Opinion mining of an educational resource

4.2 Generating opinion mining statistics for teachers

In e-learning contexts, teachers must be aware of the scope, limitations, advantages, and disadvantages of the educational material that they employ. This allows them to provide and develop meaningful educational resources and thus satisfy student learning needs and interests. In this sense, student opinions allow teachers to know the efficiency of their teaching methods and materials and determine whether improvement changes are necessary. The following section thoroughly describes the interaction process between our platform and a teacher as the user.

1. Once the teacher has logged in, the system displays a graphical interface, as illustrated in Figure 8, where the teacher can visualize the educational resources that he/she has developed. Similarly, the interface highlights the prominent polarity of each resource. That is, the polarity degree with more reviews. Green is used to highlight positive polarity, red for negative polarity, and neutral for gray polarity.

2. Any resource displayed on the interface can be selected to obtain additional details on its polarity. Figure 9 shows resource number 286 being selected, and according to the graph, 66.7% of its reviews are negative, 16.7% are positive, and 16.7% are neutral.

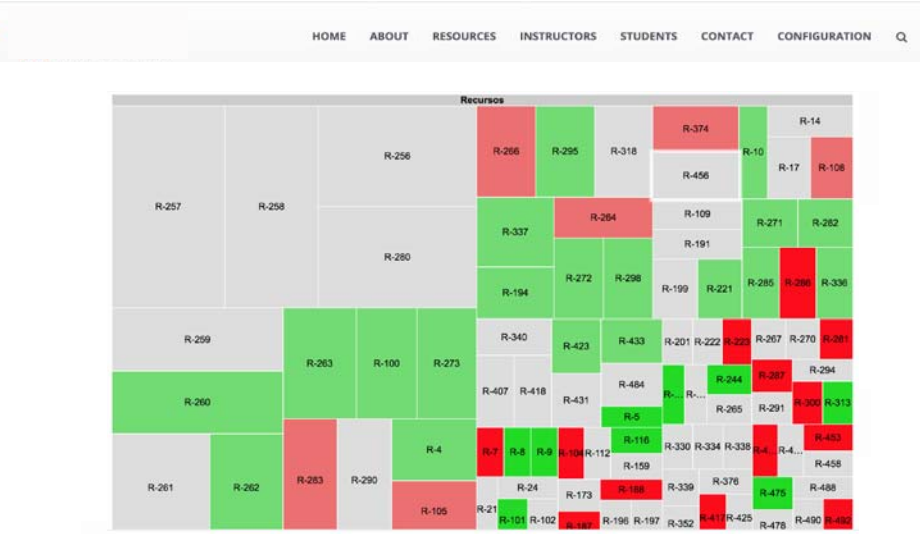

Figure 8: User resources and resource polarity.

Figure 9: Polarity of a specific resource.

5 Experiments

5.1 Data

The experiments conducted to test the platform involved the use of a labelled corpus as the context of the educational resources. Due to a lack of available corpora and datasets in Spanish for this domain, we provided the platform with our own collection of texts (see section 3.1 Data collection). The corpus generation process can be described as follows:

1. For six months, a group of students used the platform and commented on the educational resources that were previously created by teachers.

2. Duplicate comments were removed by means of an automatic filter.

3. Two experts in the education domain analyzed the comments and discarded those that did not denote an opinion.

4. Every comment was manually reviewed and labelled by the experts in terms of its polarity: positive, negative, or neutral. In total, 1,378 opinions were collected: 520 positive, 455 negative, and 403 neutral.

5. We use inter-annotator agreement measure to ensure consistent annotations. The agreement calculated at this stage using the Cohen's κ score was satisfactory with a $\kappa = 0.67$.

6. In order to ensure a balanced corpus, only 400 random opinions from each polarity class were selected. The final corpus therefore contained 1,200 opinions.

5.2 Evaluation and results

Our machine-learning-based opinion mining system was assessed with cross-validation, which is an evaluation technique that involves reducing the dependency ratio between the training data and the testing data. The data set is partitioned into k subsets, using k-1 partitions to build the model and one to perform the evaluation. The process is repeated k times; then, at each iteration, the evaluation subset is replaced by one of the other partitions. We used a ten-fold cross-validation [Martínez Cámara 2016]. Thus, for each iteration, 1,080 out of the 1,200 opinions were used to train the algorithm, whereas the remaining 120 were used for the evaluation process. Then, to assess the performance of our system, the precision, recall, and the F-measure metrics were used. These metrics were proposed by [Salton and McGill 1983] and are commonly employed to validate text classification systems, including opinion mining systems. Precision represents the proportion of predicted positive cases that are real positives (see Eq. 1), whereas recall is the proportion of actual positive cases that were correctly predicted as such (see Eq. 2). Finally, the F-measure is the harmonic mean of precision and recall (see Eq. 3).

$$\boldsymbol{Precision} = \frac{\boldsymbol{True\ positives}}{\boldsymbol{True\ positives + False\ positives}} \quad \text{Eq. 1}$$

$$\boldsymbol{Recall} = \frac{\boldsymbol{True\ positives}}{\boldsymbol{True\ positives + False\ negatives}} \quad \text{Eq. 2}$$

$$\boldsymbol{F - measure = 2} * \frac{\boldsymbol{Precision * Recall}}{\boldsymbol{Precision + Recall}} \quad \text{Eq. 3}$$

In a multiclass classification, precision, recall, and the F-measure are calculated for each class (i.e. positive, negative, and neutral). Therefore, to generate an overall evaluation of our system, the evaluation results from each class were combined. To this end, we applied the macroavering metric [Lewis 1992], which is the arithmetic mean of precision, recall, and the F-measure, where the quotient is the number of classes used in the prediction. In this sense, the Macro-Precision and Macro-Recall equations can be proposed as follows:

$$\boldsymbol{Macro - Precision} = \frac{\sum_{i=1}^{|C|} \boldsymbol{Precision}}{|C|} \qquad \text{Eq. 4}$$

$$\boldsymbol{Macro - Recall} = \frac{\sum_{i=1}^{|C|} \boldsymbol{Recall}}{|C|} \qquad \text{Eq. 5}$$

The macro average F-score is the harmonic mean of the macro-precision and macro-recall scores. The metrics presented before were obtained from a confusion matrix. Table 3 below summarizes the results for precision, recall, and the F-measure. The first column lists the number of iterations run, whereas the following columns list the scores obtained for each class, as well as the system's overall performance score.

		Precision	**Recall**	**F-Measure**
IT1	Positive	0.8057	0.8500	0.8273
	Negative	0.8571	0.9000	0.8780
	Neutral	0.8101	0.7250	0.7652
	Macroavering	**0.8243**	**0.8250**	**0.8235**
IT2	Positive	0.7972	0.8650	0.8297
	Negative	0.8349	0.8850	0.8592
	Neutral	0.8187	0.7000	0.7547
	Macroavering	**0.8170**	**0.8167**	**0.8146**
IT3	Positive	0.8038	0.8400	0.8215
	Negative	0.8243	0.9150	0.8673
	Neutral	0.8166	0.6900	0.7480
	Macroavering	**0.8149**	**0.8150**	**0.8123**
IT4	Positive	0.7887	0.8400	0.8136
	Negative	0.8037	0.8800	0.8401
	Neutral	0.7976	0.6700	0.7283
	Macroavering	**0.7967**	**0.7967**	**0.7940**
IT5	Positive	0.7792	0.9000	0.8353
	Negative	0.8416	0.8500	0.8458
	Neutral	0.8263	0.6900	0.7520
	Macroavering	**0.8157**	**0.8133**	**0.8110**

IT6	Positive	0.8119	0.8850	0.8469
	Negative	0.7939	0.9050	0.8458
	Neutral	0.8312	0.6400	0.7232
	Macroavering	**0.8123**	**0.8100**	**0.8053**
IT7	Positive	0.8244	0.8450	0.8346
	Negative	0.7939	0.9050	0.8458
	Neutral	0.8024	0.6700	0.7302
	Macroavering	**0.8069**	**0.8067**	**0.8035**
IT8	Positive	0.8161	0.9100	0.8605
	Negative	0.8491	0.9000	0.8738
	Neutral	0.8424	0.6950	0.7616
	Macroavering	**0.8359**	**0.8350**	**0.8320**
IT9	Positive	0.8157	0.8850	0.8489
	Negative	0.8531	0.9000	0.8759
	Neutral	0.8430	0.7250	0.7796
	Macroavering	**0.8373**	**0.8367**	**0.8348**
IT10	Positive	0.8213	0.8500	0.8354
	Negative	0.8372	0.9000	0.8675
	Neutral	0.8146	0.7250	0.7672
	Macroavering	**0.8244**	**0.8250**	**0.8233**
AVG		0.8185	0.8180	0.8154

Table 3: Evaluation results

As Table 3 indicates, the system's average scores for precision, recall, and the F-measure are 81.85%, 81.80%, and 81.54%, respectively. Such results are encouraging, as they demonstrate that the system can successfully detect the polarity of educational resource reviews written in Spanish. As for the iterations, "IT9" achieved the best results with a precision score of 83.73%, a recall score of 83.67%, and an F-measure score of 83.48%. Conversely, iteration "IT4" showed the least positive results: 79.67% for precision, 79.67% for recall, and 79.40% for the F-measure. Finally, the neutral class had the least favorable results among the three classes, thereby implying that neutral opinions are a challenge for our opinion mining system. In fact, it is usually difficult to distinguish between neutral sentiment and non-sentiment bearing sentences.

5.2.1 Comparison with related work

We have carried out a comparative analysis of related proposals on opinion mining research in the educational domain. In particular, we have considered reported Precision, Recall, F-measure and Accuracy scores for that purpose. According to the results obtained, which are shown in Table 4, the majority of the related proposals are focused on the English language. We believe that the interest in the English language is due to the fact that it is an official language in many countries, and most of the content on the Internet is written in this language. Table 4 also shows that our proposal achieved slightly higher Precision, Recall and F-measure scores than practically any of related works. Only one of the works that is focused on the English language obtained better results than our proposal. As regards the Spanish language, related works have been evaluated using different standard metrics, thus making it difficult to determine whether a work is better or worse than other.

Comparing different opinion mining approaches may be difficult for several reasons. In this work, we have found three of them: 1) the proposals presented in [Kravvaris and Kermanidis 2017], [Rajput et al. 2016], [Kechaou et al. 2011], [Dhanalakshmi et al. 2016], [S. Rani and Kumar 2017], and [Zarra et al. 2016] are focused on a language other than Spanish, 2) the corpora used for each experiment significantly differ in content, size, and language; a fair comparison of two opinion mining methods would require the usage of the same testing corpus and 3) the corpora used by the works are not publicly available in all the cases.

	Language	**Precision**	**Recall**	**F-measure**	**Accuracy**
[Kravvaris and Kermanidis 2017]	English	--	--	--	82.25
[Rajput et al. 2016]	English	75.00	82.00	73.00	
[Kechaou et al. 2011]	English	80.00	80.00	79.90	---
[Ortigosa et al. 2014]	Spanish	--	--	--	83.27
[Dhanalakshmi et al. 2016]	English	99.75	97.07	--	99.11
[Zarra et al. 2016]	English	92.81	60.32	73.12	63.18
Our proposal	Spanish	81.85	81.80	81.54	--

Table 4: Results of the comparative analysis (evaluation).

6 Conclusions and Future Work

This paper introduces an educational platform in Spanish that relies on machine-learning-based opinion mining techniques to rate and comment on educational resources. In this sense, the system allows students and teachers to structure their own e-learning/e-teaching experiences by selecting those educational resources that match their interests and meet their needs. On the one hand, our opinion mining approach enables students to select those resources that, according to the reviews, are likely to meet their needs and satisfy their learning expectations. On the other hand, teachers can rely on such reviews to determine the actual impact of their educational resources on student learning. Finally, the system is also able to generate personalized profiles.

To assess the performance of our system, we conducted a set of experiments on a corpus of 1,200 resource opinions (400 positive, 400 negative, and 400 neutral). The system obtained encouraging performance results, with average scores of 81.85% for precision, 81.80% for recall, and 81.54% for the F-measure. Similarly, our findings suggest that the classification of neutral-sentiment bearing content is still a challenge for our system.

For future work, we have planned to perform opinion mining at feature level, which will allow identifying and analyzing specific aspects of learners' comments. This will finally allow teachers to know the specific aspects that they should improve in the teaching processes. We have also planned to use recommendation techniques to provide the system with the ability to suggest educational resources based on ontology-based user profiles and user preferences. In this context, we will explore the possibility of creating semantic profiles also for educational resources. Furthermore, it would be interesting to increase the system's information sources by taking into consideration social networks other than Facebook® and Twitter®, such as YouTube®, SlideShare®, Scribd®, and Picasa®, among others. As regards evaluation, we have planned to experiment with corpus of different sizes to determine the impacts on performance metrics. Likewise, we are going to carry out a hyperparameter optimization with the objective of identifying the best algorithm and configuration for our problem. Finally, a comparative analysis with a lexicon-based opinion mining approach could help us to identify some strengths and weaknesses of the proposed opinion mining approach.

Acknowledgements

The authors are grateful to the National Technological Institute of Mexico (Tecnológico Nacional de Mexico, TecNM) for supporting this work. The research was also sponsored by Mexico's National Council of Science and Technology (CONACYT) and the Secretariat of Public Education (SEP) through the PRODEP program.

References

[Alimam, Seghiouer and Elyusufi 2014] Alimam, M. A., Seghiouer, H., Elyusufi, Y.: 'Building profiles based on ontology for career recommendation in E-learning context'; In 2014 International Conference on Multimedia Computing and Systems (ICMCS) (2014), pp. 558–562. https://doi.org/10.1109/ICMCS.2014.6911346

[Anido et al. 2002] Anido, L. E., Fernández, M. J., Caeiro, M., Santos, J. M., Rodríguez, J. S., Llamas, M.: 'Educational metadata and brokerage for learning resources'; Computers & Education, Vol. 38, No. 4 (2002), pp. 351–374. https://doi.org/10.1016/S0360-1315(02)00018-0

[Auer et al. 2007] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: 'DBpedia: A Nucleus for a Web of Open Data'; In K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, et al. (Eds.), The Semantic Web SE - 52 (Vol. 4825). Springer Berlin Heidelberg (2007), pp. 722–735. https://doi.org/10.1007/978-3-540-76298-0_52

[Baccianella, Esuli and Sebastiani 2010] Baccianella, S., Esuli, A., Sebastiani, F.: 'SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.'; In LREC (Vol. 10) (2010), pp. 2200–2204.

[Brickley and Miller 2000] Brickley, an, Miller, L.: 'FOAF Vocabulary Specification'; (2000).

[Dhanalakshmi, Bino and Saravanan 2016] Dhanalakshmi, V., Bino, D., Saravanan, A. M.: 'Opinion mining from student feedback data using supervised learning algorithms'; In 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC) (2016), pp. 1–5. https://doi.org/10.1109/ICBDSC.2016.7460390

[Facebook 2018] Facebook: 'API Graph'; (2018).

[Google 2018] Google: 'Natural Language API'; (2018).

[Heiyanthuduwage, Schwitter and Orgun 2016] Heiyanthuduwage, S. R., Schwitter, R., Orgun, M. A.: 'A learning ontology with metadata and user profiles for enhancing accessibility of resources'; In 2016 IEEE Conference on e-Learning, e-Management and e-Services (IC3e) (2016), pp. 85–90. https://doi.org/10.1109/IC3e.2016.8009045

[Jansen 2014] Jansen, E.: 'NetLingo: The Largest List of Chat Acronyms and Text Shorthand'; NetLingo Inc. (2014).

[Kechaou, Ammar and Alimi 2011] Kechaou, Z., Ammar, M. Ben, Alimi, A. M.: 'Improving e-learning with sentiment analysis of users' opinions'; In 2011 IEEE Global Engineering Education Conference (EDUCON) (2011), pp. 1032–1038. https://doi.org/10.1109/EDUCON.2011.5773275

[Kravvaris and Kermanidis 2017] Kravvaris, D., Kermanidis, K. L.: 'Opinion Mining for Educational Video Lectures'; Advances in Experimental Medicine and Biology, Vol. 989 (2017), pp. 235–243. https://doi.org/10.1007/978-3-319-57348-9_20

[Leong, Lee and Mak 2012] Leong, C. K., Lee, Y. H., Mak, W. K.: 'Mining sentiments in SMS texts for teaching evaluation'; Expert Systems with Applications, Vol. 39, No. 3 (2012), pp. 2584–2589. https://doi.org/10.1016/j.eswa.2011.08.113

[Lewis 1992] Lewis, D. D.: 'Representation and Learning in Information Retrieval'; University of Massachusetts, Amherst, MA, USA (1992).

[Liu 2015] Liu, B.: 'Sentiment Analysis: Mining Opinions, Sentiments, and Emotions'; Cambridge University Press (2015).

[Martínez Cámara 2016] Martínez Cámara, E.: 'Análisis de Opiniones en Español'; (2016).

[Németh 2005] Németh, L.: 'Hunspell'; (2005).

[Ortigosa, Martín and Carro 2014] Ortigosa, A., Martín, J. M., Carro, R. M.: 'Sentiment analysis in Facebook and its application to e-learning'; Computers in Human Behavior, Vol. 31 (2014), pp. 527–541. https://doi.org/10.1016/j.chb.2013.05.024

[Rajput, Haider and Ghani 2016] Rajput, Q., Haider, S., Ghani, S.: 'Lexicon-Based Sentiment Analysis of Teachers' Evaluation' [Research article]; (2016).

[M. Rani, Nayak and Vyas 2015] Rani, M., Nayak, R., Vyas, O. P.: 'An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage'; Knowledge-Based Systems, Vol. 90 (2015), pp. 33–48. https://doi.org/10.1016/j.knosys.2015.10.002

[S. Rani and Kumar 2017] Rani, S., Kumar, P.: ‘A Sentiment Analysis System to Improve Teaching and Learning’; Computer, Vol. 50, No. 5 (2017), pp. 36–43. https://doi.org/10.1109/MC.2017.133

[Salas-Zárate, Paredes-Valverde, Rodriguez-García, Valencia-García and Alor-Hernández 2017] Salas-Zárate, M. del P., Paredes-Valverde, M. A., Rodriguez-García, M. Á., Valencia-García, R., Alor-Hernández, G.: ‘Automatic detection of satire in Twitter: A psycholinguistic-based approach’; Knowledge-Based Systems, Vol. 128 (2017), pp. 20–33. https://doi.org/10.1016/j.knosys.2017.04.009

[Salas-Zárate María del Pilar, Paredes-Valverde Mario Andrés, Limon-Romero Jorge, Tlapa Diego and Baez Lopez Yolanda 2016] Salas-Zárate María del Pilar, Paredes-Valverde Mario Andrés, Limon-Romero Jorge, Tlapa Diego, Baez Lopez Yolanda: ‘Sentiment Classification of Spanish Reviews: An Approach based on Feature Selection and Machine Learning Methods’; Journal of Universal Computer Science, Vol. 22, No. 5 (2016), pp. 691–708. Retrieved from http://www.jucs.org/jucs_22_5/sentiment_classification_of_spanish

[Salton and McGill 1983] Salton, G., McGill, M. J.: ‘Introduction to modern information retrieval’; McGraw-Hill (1983). Retrieved from https://dl.acm.org/citation.cfm?id=576628

[Sharma and Ahuja 2016] Sharma, M., Ahuja, L.: ‘A Novel and Integrated Semantic Recommendation System for E-Learning Using Ontology’; In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies. New York, NY, USA: ACM (2016), p. 52:1–52:5. https://doi.org/10.1145/2905055.2905110

[Tarus, Niu and Yousif 2017] Tarus, J. K., Niu, Z., Yousif, A.: ‘A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining’; Future Generation Computer Systems, Vol. 72 (2017), pp. 37–48. https://doi.org/10.1016/j.future.2017.02.049

[Tewari, Saroj and Barman 2015] Tewari, A. S., Saroj, A., Barman, A. G.: ‘e-Learning Recommender System for Teachers using Opinion Mining’; In Information Science and Applications. Springer, Berlin, Heidelberg (2015), pp. 1021–1029.

[Twitter 2018] Twitter: ‘Twitter API’; (2018).

[Zarra, Chiheb, Faizi and Afia 2016] Zarra, T., Chiheb, R., Faizi, R., Afia, A. El: ‘Cloud computing and sentiment analysis in E-learning systems’; In 2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech) (2016), pp. 171–176. https://doi.org/10.1109/CloudTech.2016.7847695