Extraction and recommendation of experts on topics of interest in social networks as an educational tool

Paúl E. Vintimilla-Tapia, Jack F. Bravo-Torres
Grupo de Investigación GITEL
Universidad Politécnica Salesiana
Cuenca, Ecuador
Email: {pvintimilla, jbravo}@ups.edu.ec

Karina de Lourdes Serrano-Paredes¹, Isabel C. Mesa-Cano²
¹Facultad de Medicina, Universidad Católica de Cuenca
Cuenca, Ecuador
²Facultad de Enfermería, Universidad Católica de Cuenca
Cuenca, Ecuador
Email: {kserrano, imesac}@ucacue.edu.ec

Abstract—Nowadays, Information and Communication Technologies (ICTs), along with Web 2.0 technologies, are enabling the globalization of Internet, providing a mean of access for the creation, dissemination and discussion of content. In this way, anyone can consult information of interest, assimilate it and turn it into useful knowledge. However, in recent years this new technological trend has driven users to generate a large amount of personal content, leaving aside any quality index, which translates into a new social problem known as information overload, infodication or infobesity. The contact with information that is constantly increasing and of which validity has not been proven can cause difficulties, from the assimilation of knowledge to psychological disorders (anguish). The educational field is no stranger to this situation, as students use technology to support their academic processes. This research proposes the development of an experts recommendation tool (individuals who significantly manage a topic of interest) based on Twitter and Mendeley with a semi-supervised approach. In a Web application, keywords related to a topic of interest are entered and extracted from potential Mendeley experts, and then their accounts are located on Twitter. With this information, a user validates whether the Twitter profiles correspond to experts and authorizes the publication of a recommendation to students. With the semi-supervised approach, the accuracy of the recommendations is 100%, so the results obtained are promising.

Keywords—Expert recommendation, information overload, Mendeley, social networking, Twitter.

I. INTRODUCTION

Today, due to the constant development of the Internet, humanity is experiencing a new social paradigm, in which access to information is globalized, allowing the extraction of useful knowledge [1]. The so-called knowledge society is born based on two technological currents: Information and Communication Technologies (ICTs) and Web 2.0 technologies, which although they present differences at a conceptual level, constitute the support of modern services and applications presented to users [2]. On the one hand, the first one groups different elements and techniques that are used both for data transmission and processing and for reducing digital divides [3]. On the other hand, the second marks an evolution of the functionalities of the World Wide Web to give users greater prominence in the creation, dissemination and discussion of content [4]. As can be seen, ICTs are directly related to the way information is accessed, while Web 2.0 provides the tools and resources needed to obtain useful knowledge [5].

Over the last few years, Web 2.0 technologies have been defined from different perspectives by several authors [4, 6], establishing a general concept that outlines the use of the Web to enable users to collaborate with each other, actively engage in content creation, generate knowledge, and share information online [7]. New tools and resources such as blogs, microblogs, wikis, RSS syndicators, tag-based folksonomies, social markers, multimedia sharing and social networking sites are emerging, which encourage social behaviour with a participatory and narrative approach, in which information is shared through text, images, audio and video [8]. Faced with such a situation, a wide range of educational possibilities is deployed, since, by allowing the social participation of a group of individuals in the elaboration of content, the individualized thinking that can be found in the classroom is avoided [9]. In addition, the centralized teaching processes in the educator, move towards a centralized model in the student that breaks down the spatial-temporal barriers, dependent on a physical environment. The student becomes the main actor in his or her education and the figure of the teacher becomes a mediator of knowledge, whose main objective is to lay the foundations for learning to learn, motivating the need for lifelong learning [10].

In this sense, one of the main problems that Web 2.0 seeks to solve is the poor social interaction that develops within the academy [11]. Specifically with the deployment of social networks, communication flows more naturally and students feel more comfortable interacting with their peers and teachers, forming a virtual educational community [12]. In this way, a new theory of learning, known as connectivism, emerges, which takes a different starting point from other theories (see TABLE I) [13]–[15], focusing on helping the student to be autonomous, independent and self-taught, rather than covering the contents to be taught. A partnership is created with critical and thoughtful individuals who are able to control their own learning process, specialize in what they want to investigate and learn from the knowledge their peers manage [16]. Their main characteristics are the transfer of education from the classrooms to a personal space, the elimination of strict
The factors that motivate their use are based on three social theories: joint intention, social influence and social presence. On the one hand, the theory of joint intention sets out the commitment to collaborate in the implementation of a given action in conjunction with a predefined group of participants, understood as an agreement expressed in terms of “Together we will develop X (X symbolizes a joint action)” [20]. Thus, in a group act, each individual sees himself or herself as part of a social representation and changes his or her individualized behavior [23]. On the other hand, the theory of social influence determines the attitudinal changes produced by different factors that occur when using a social network [24]. Among the most notorious are: (i) compliance, which occurs when one seeks to motivate a particular behaviour in return for a reward; (ii) internalization, which refers to the adoption of common objectives to achieve idealized goals; and (iii) identification, which occurs when establishing or maintaining contact with like-minded people [25]. Finally, the theory of social presence highlights the degree of personal promise achieved during satisfactory social interactions, and hence the prominence of interpersonal relationships [26]. This is reflected in the importance of the presence of several individuals in the same virtual environment, which guarantees the existence of human contact through methods that produce feelings similar to face-to-face communication [27].

From an academic point of view, social networks are being significantly introduced into learning/teaching processes, due in large part to the nature in which new generations are developing [28]. Students use them as an ideal space to exchange information, which later becomes knowledge, in a fast, simple and comfortable way, giving teachers an invaluable opportunity to improve their techniques [20]. This new trend is based on the informality found in a large number of social networks, since in their beginnings they were used as a means of leisure and entertainment [29]. However, over time, they have been found to provide a number of alternatives for addressing current educational challenges, which aim to develop individuals with self-learning, critical thinking, collaborative working and research skills [30]. They are also involved in locating experts on different topics of interest, providing a valuable resource for knowledge transfer [31]. All these peculiarities make it possible to deploy several applications [32]:

- Subjects networks: In some cases, a subject-specific net-
work is created for the purpose of establishing dialogues, consulting doubts, and even carrying out tasks. In order to take advantage of its social capacities, it is recommended that it be made up of a significant number of participants.

- Internal communities networks: These are established by each educational institution with the purpose of creating a sense of belonging to a real community in both students and teachers. Within it, different groups can be deployed, according to specific objectives [33]:

  - Consultation groups: They provide a private virtual space for students and teachers of the same subject to get in touch. Their main difference with the subjects networks is that they are linked within a larger network, which means that they can accommodate a small number of participants without wasting their performance. Significant examples of use include homework consultations and grades review.

  - Informative groups: They are an appropriate space for teachers to place instructions on the tasks that students must perform and general indications of each subject.

  - Students groups: When developing activities within a collaborative group, it is necessary to provide students with a private space for the exchange of ideas and organization of tasks to meet their objectives.

- Tutoring networks: Their main objective is to offer the knowledge of experts to solve doubts and concerns about any subject of study. Using tools such as question/answer platforms and blogs, students resort to the expertise of a known or unknown user to reinforce their learning.

In summary, education is nothing more than a sophisticated process of information exchange that seeks to train each individual in subjects of his or her interest, offering characteristics to achieve correct development throughout his or her life [34]. This statement justifies the great reception that Web 2.0, specifically social networks, have had as valid resources for the optimization of learning/teaching processes, successfully enhancing the experience of students and teachers [35]. Nevertheless, a significant problem is emerging, as a product of the free creation and dissemination of content, known as information overload, infoxication or infobesity [36]–[38]. Any user has the power to originate personal content, but of the millions of data that go through the Internet, only a small portion has scientific backing, importance and usefulness [39]. Thus, the human brain, being a limited cognitive resource, does not have the processing power to handle the exponential increase in information displayed on the Web [40]. It is also difficult to locate users who are considered experts [41].

A solution to this difficulty has been found in the design of tools for filtering, selecting and choosing useful content, known as recommender systems, which are responsible for suggesting the start of a given action [42]. As an educational resource, they are adapted to the needs, tastes and preferences of each student, becoming a guide to the activities they can develop, including interpersonal relationships with other users (experts) [43], [44]. With regard to this last point, the bibliography includes several important documents, such as [45] which mentions that question/answer communities (QACs) are valuable information resources that provide a platform for knowledge sharing. Finding potential experts within a QAC is beneficial for solving problems resulting of the low participation rate, the long response time and the low quality of the content. The evolution of personal experience over time is then taken into account in order to identify potential experts for the future, generating the design of a classification framework. The experimental results indicate the efficiency of the proposed model in comparison with traditional techniques, obtaining an improvement of 39.7%.

In [46], greater interest is shown in finding groups or teams rather than individual entities. In this way, a new localization methodology is introduced that is based on the detection of useful knowledge from a repository of heterogeneous documents. Thus, experience is added to the members of the same group based on the resolution of a task, to extract personal documents associated with each one and determine how closely they are linked to a topic of interest, evaluating the degree of group knowledge. The experimental results show high absolute scores in terms of mastery of a particular topic by each group.

Finally, in [47] the authors state that with the growing demands for knowledge, users often rely on virtual communities for the exchange of information. In some cases, the content generated is of low quality, preventing the development of an effective management system that facilitates the search for experts. In addition, existing systems assess each user’s experience based on the content of their scientific publications or on their social status within a research community, with very few studies considering both aspects. Consequently, an expert search algorithm, known as ExpertRank, is designed to evaluate the experience based on both publication relevance and authority within his or her community.
Expert detection is a classification task, most of which is carried out using unsupervised artificial intelligence algorithms. This means that no human factor is incorporated to validate the information extracted before recommending it to an user, and although promising results are obtained, efficiency is not optimal. This document proposes a semi-supervised approach to the design of an expert recommendation tool, in which a teacher enters keywords on topics of interest according to the subjects he or she teaches. With Mendeley’s help, profiles of potential experts are extracted and their accounts are located on Twitter for subsequent recommendation to students. The following section explains the design and operation of the recommendation tool.

III. DESIGN OF THE EDUCATIONAL TOOL FOR THE RECOMMENDATION OF EXPERTS

The Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia\(^1\) (CEDIA), is an entity that promotes the exploration of innovative solutions by linking Ecuadorian institutions. To this end, it brings together researchers, teachers and students through projects, competitions and scientific progress initiatives, creating a circle of constant growth that translates into collaborative work to generate and strengthen research networks in national and international spheres, to the benefit of society. Its main objective is to lead the research community in academia and Ecuadorian business, offering several tools and resources, high-speed Internet connections, open virtual mass courses (MOOC) and countless certifications.

The majority of CEDIA’s staff is made up of teachers who are aware of the difficulties that arise day by day in the students’ lives, highlighting the negative effects of an overload of information due to their constant exposure to the Internet. Then, taking advantage of the communicative and inter-relational perks that are included in Web 2.0 technologies, specifically in social networks, the development of an academic tool was proposed with the purpose of searching and recommending experts in different topics of interest. Three entities are working together: Mendeley, which, although it is an academic repository, can be seen as a social network of researchers; Twitter, which is a social network of microblogging; and users, most of whom are expected to be teachers, achieving a semi-supervised approach that guarantees 100% accuracy in the final recommendations. The use of five modules is therefore necessary:

- **User module**: intervenes in the correct interaction between a user and the experts recommendation tool. It is no more than a responsive web application\(^2\) provided with the necessary elements to enter keywords related to topics of interest, visualize the experts that have been located in Mendeley, validate whether the Twitter accounts extracted correspond to each expert and schedule a date for the publication of the recommendation.

- **Experts extraction module**: Its action is supported by the functionalities offered by certain libraries, commonly known as APIs (Application Programming Interface), which deliver a set of subroutines, functions and procedures, which are used as an abstraction layer within a developing software. The expert recommender requires both the Mendeley API and the Twitter API to work. This is due to the fact that, on the one hand, possible experts on a given topic are searched, making it necessary to interact and analyze exhaustively several scientific articles, abstracts and bibliographic references. On the other hand, to locate selected Mendeley experts on Twitter, it is essential to extract a certain number of tweets from candidate accounts to apply a natural language processing algorithm and determine any existing relationships.

- **Experts validation module**: Once potential experts have been located on a topic of interest at Mendeley and their Twitter accounts have been found, all options are stored in the database to present who requested the search. At this point, each user analyzes the Twitter accounts in detail, as there may be problems related to ambiguities, and those that are considered optimal, in the sense of the information they share, are selected for publication. The remaining accounts are discarded and deleted from the database.

- **Experts publication and recommendation module**: With the identification of the experts’ accounts, the date (day and time) for the publication of a recommendation is selected. To do this, a tweet is created that includes the name of the expert, his or her work area and Twitter account. The publication of the tweet is done through a private account, created only for this purpose and followed by students of interest, avoiding a leisure and information pollution environment.

It is important to mention that this tool was designed with the aim of allowing teachers to recommend students about the experts they can follow. For this reason, although the experts recommender is within reach of anyone, it is convenient to have an adequate handling on the topics to be worked with. The following subsection describes how it works.

A. Operation

To recommend experts on a topic of interest, the tool starts working when a user logs in to the web application; if not registered, a new account can be created by validating an email address and password. Once inside, a set of keywords on a topic of interest are entered, which are recommended to be specific on a particular theme, avoiding any ambiguity. Based

---

1. https://www.cedia.edu.ec/
2. It is a web design technique that allows you to create applications that adapt to the browser in which they run, showing an optimized version.
on these, a search for publications is conducted within Mendeley. For each document found, it is consulted about its authors and their interactions with their peers (cited authors), creating a directed graph (digraph) of author/follower relations. At this point, the documents obtained in the initial search are analyzed again to extract their abstracts and construct a histogram of lemmas’ lengths (Mendeley Lemas), which is based on the fact that longer words have more meaning, and are therefore more common within an expert’s vocabulary. This procedure will allow to filter out unwanted Twitter accounts. With the expert candidates identified in Mendeley, the PageRank algorithm is applied to the target graph to filter out the authors who have the most followers among their peers (5 by default). Thus, the Twitter API is used to search for all accounts that match them. For each account the last 200 tweets posted in English are gotten, to create another histogram of lemmas’ lengths (Twitter Lemas). It verifies if there are pairs of Mendeley Lemas – Twitter Lemas. In the case of a match, it is concluded that the Twitter account belongs to the author of Mendeley, so it will be registered in the database as the account of a potential expert. When all accounts have been registered, the tool will notify the user who initiated the search via email to validate the information. Finally, the user selects the accounts that really belong to experts for publication and recommendation by means of a tweet; the remaining accounts are discarded. The above steps can be seen in the diagram in Fig. 2.

IV. RESULTS OF A FIRST IMPLEMENTATION

For a first implementation, the resources provided by an Intel Xeon server with 16 GB of ram and Ubuntu Server 16.06 operating system were available. Here all the services and complementary packages that the tool needs to deploy its correct operation were installed. In addition, two operation tests were conducted to verify the validity of the recommendations. On the one hand, experts linked to the topic “Artificial Intelligence” were searched using the keywords “Bayes Perceptron”. On the other hand, experts related to the topic “Surgery” were found through the keywords “Ergonomics laparoscopic surgery”. The results achieved before the user validation are presented in the TABLE II.

As can be seen, in the first case, 5 accounts of potential experts were found on Twitter. However, one of them talks about “Marketing and finance”, which has nothing to do with the search topic that was “Artificial Intelligence”. In this particular account, the reason for the erroneous recommendation is due to the algorithm for detecting expert accounts on Twitter. When handling a validation by comparing lemas based on the length of keywords between the abstracts of an author’s documents and a user’s tweets, confusion can occur due to the vocabulary used. This is known as expert ambiguity, reflecting the fact that he found a Twitter account that shares his or her name with a Mendeley author, but is linked to a different field of interest. In the second case, 4 expert Twitter accounts were found, all linked to the search topic. The accuracy of the algorithm in this search topic is justified by the unique characteristics of the lexicon developed by individuals related to the field of medicine. On the other hand, a factor that has a significant influence on a correct search for experts is the

Once understood how the expert recommender works, in the following section two cases of use and their respective analysis are exposed.
TABLE II
EXTRACTION OF POTENTIAL EXPERTS PRIOR TO USER VALIDATION

<table>
<thead>
<tr>
<th>Topic</th>
<th>Keywords</th>
<th>Twitter account</th>
<th>Account detail summary</th>
<th>Relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial Intelligence</td>
<td>Bayes Perceptron</td>
<td>@hjelmj</td>
<td>Electrochemist with an interest in scientific computation</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@aspios74</td>
<td>PhD in Computer Science. Interested in pattern recognition and computer vision</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@caamitkulki</td>
<td>MBA in marketing with love of finance</td>
<td>NO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@peterajohnson</td>
<td>Lover of geospatial technology, Assistant professor at the University of Waterloo</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@jeridfrancom</td>
<td>Cognitive scientist, linguist and &quot;data geek&quot;</td>
<td>YES</td>
</tr>
<tr>
<td>Surgery</td>
<td>Ergonomics laparoscopic surgery</td>
<td>@MrRhinoplasty</td>
<td>Surgeon specializing in rhinoplasty</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@thomasleemd</td>
<td>Orthopedic surgeon. Special interest in the foot/ankle</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@drojasz</td>
<td>Neurosurgeon. Institute of Neurosurgery. Academic at the Universidad de Chile</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td></td>
<td>@Dr_D_Rosenberg</td>
<td>Plastic surgeon</td>
<td>YES</td>
</tr>
</tbody>
</table>

keywords with which the recommender works. If keywords that are too general are specified, accounts could be extracted from users who have nothing to do with a topic of interest. Finally, it should be noted that finding a Twitter expert’s account does not ensure that they are constantly publishing useful content.

In this regard, if the recommender’s approach had been unsupervised, approximately 90% accuracy would have been achieved by the ambiguity of experts, but it would not be possible to ensure that the content published in each account is closely linked to a particular topic. For these reasons, by implementing the help of a connoisseur teacher through a semi-supervised approach, a more robust filtering is developed that helps to discard erroneous accounts, reaching 100% accuracy in the recommendations made. This statement supports the limitation that, although the application can be used by any user, it is advised that the recommendations be deployed by individuals with established knowledge of the field of study with which they will work.

V. CONCLUSIONS AND FUTURE WORK

The information overload is a serious problem that compromises the assimilation of knowledge that occurs on the Internet every day. Any user can generate content on their topics of interest, but no one can guarantee that the information they are trying to transmit is valid or invalid. In such a situation, it is of vital importance to distinguish users with established knowledge and experience that allows them to be considered as experts, to bring them closer to the students. For this reason, this paper presented the design of an experts recommender based on Mendeley and Twitter with a semi-supervised approach. Although the results are promising, due in large part to the human validation of expert accounts prior to the recommendation, it should be borne in mind that their effectiveness depends on how specific the topic is, since very general words draw experts from other areas. Also, it may be the case that the expert does not have a Twitter account or that the tool confuses it by falling into an ambiguity of experts, since the correlation of lemas only distinguishes if the writing style corresponds to an expert.

As future work, it is proposed to use other research networks as question/answer communities to optimize the expert search algorithm through a non-supervised approach.

ACKNOWLEDGMENT

The authors extend their sincere gratitude to the Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia, CEDIA, for the funding provided for research, development and innovation, by means of the projects generated through CEDIA’s Telemedicine/Telehealth Working Group.

REFERENCES
