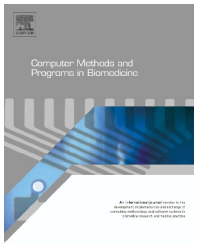
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE XXX (2015) XXX–XXX

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Automatic extraction and identification of users' responses in Facebook medical quizzes

Alejandro Rodríguez-González[a,*,1], *Ernestina Menasalvas Ruiz*[b,1], *Miguel A. Mayer Pujadas*[c,1]

[a] *ETS de Ingenieros Informáticos, Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, 28660 Madrid, Spain*

[b] *Centro de Tecnología Biomédica, Universidad Politécnica de Madrid, Campus de Montegancedo, Pozuelo de Alarcon, 28223 Madrid, Spain*

[c] *Research Programme on Biomedical Informatics, Hospital del Mar Medical Research Institute (IMIM) – Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona (PRBB), Dr. Aiguader, 8808003 Barcelona, Spain*

ARTICLE INFO

Article history:
Received 24 October 2015
Received in revised form
5 December 2015
Accepted 18 December 2015

Keywords:
Facebook
Health-related websites
Data retrieval
Medical quizzes

ABSTRACT

Background: In the last few years the use of social media in medicine has grown exponentially, providing a new area of research based on the analysis and use of Web 2.0 capabilities. In addition, the use of social media in medical education is a subject of particular interest which has been addressed in several studies. One example of this application is the medical quizzes of The New England Journal of Medicine (NEJM) that regularly publishes a set of questions through their Facebook timeline.

Objective: We present an approach for the automatic extraction of medical quizzes and their associated answers on a Facebook platform by means of a set of computer-based methods and algorithms.

Methods: We have developed a tool for the extraction and analysis of medical quizzes stored on Facebook timeline at the NEJM Facebook page, based on a set of computer-based methods and algorithms using Java. The system is divided into two main modules: Crawler and Data retrieval.

Results: The system was launched on December 31, 2014 and crawled through a total of 3004 valid posts and 200,081 valid comments. The first post was dated on July 23, 2009 and the last one on December 30, 2014. 285 quizzes were analyzed with 32,780 different users providing answers to the aforementioned quizzes. Of the 285 quizzes, patterns were found in 261 (91.58%). From these 261 quizzes where trends were found, we saw that users follow trends of incorrect answers in 13 quizzes and trends of correct answers in 248.

Conclusions: This tool is capable of automatically identifying the correct and wrong answers to a quiz provided on Facebook posts in a text format to a quiz, with a small rate of false negative cases and this approach could be applicable to the extraction and analysis of other sources after including some adaptations of the information on the Internet.

* *Corresponding author.*
E-mail addresses: alejandro.rodriguezg@upm.es (A. Rodríguez-González), ernestina.menasalvas@upm.es (E. Menasalvas Ruiz), mmayer@imim.es (M.A. Mayer Pujadas).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.cmpb.2015.12.025
0169-2607/

1. Introduction

In the last few years the use of social media in medicine has grown exponentially, providing a new area of research based on the analysis and use of Web 2.0 capabilities. This has affected the way that people share and exchange ideas, opinions and feelings [1]. Furthermore, it has increased the amount of information available on the Internet under the concept of user generated content (UGC) [2]. In medicine, social media has a deep impact because the number of websites with health-related information has been growing very fast in the last few years [3], making this information available to a wider audience access. However, the lack of control and quality [4] regarding this information also contributes to the generation of low quality medical information which could be dangerous for the potential users [4,5].

In addition, the use of social media in medical education is a subject of particular interest which has been addressed in several studies such as the one undertaken by the Penn State College of Medicine in relation with the use of Twitter, YouTube, Flickr, blogging and Skype to promote students learning [6]. Another interesting analysis was made by Bahner et al. [7] which studies the effectiveness of using Twitter and Facebook as an educational tool by sharing messages related to a particular subject. Cheston et al. [8] also carried out a study where the use of social media in education was analyzed.

In the context of medical education, there are several social media platforms that offer medical knowledge in different ways and formats. Twitter accounts like USMLE [9] or Crush USMLE [10] offer relevant information regarding the United States Medical Licensing Examination (USMLE). Radiology Signs [11] and Radiopaedia [12] are two Facebook pages focused either on the publication of medical images with relevant information or on asking the page users for a diagnosis on the comments of the post. Similar to those previously mentioned is the New England Journal of Medicine (NEJM) Image Challenge Application [13], where your answer to a concrete challenge can be submitted offering basic statistics compared with other answers and interactive medical cases provided by the NEJM official page [14].

Medical quizzes are another interesting category of educational content that can be found in social media platforms. Facebook pages such as Medical Quiz [15] or medical quizzes [16] provide these types of educational questions. However, the format used in these pages is heterogeneous, hindering later analysis of the questions and answers published.

Medical quizzes of The New England Journal of Medicine (NEJM) (one of the most relevant journals in medicine around the world), regularly publishes a set of questions through their Facebook timeline allowing NEJM Facebook page users to answer, discuss or just learn about the knowledge around the question. In particular, the "Medical Quiz" consists of a question about a case study that is accompanied by a set of possible answers. The community manager creates a post where the question is formulated generally with a link to some article or interesting text related to the proposed question. Fig. 1 shows a screenshot with an example of this type of quiz and some answers.

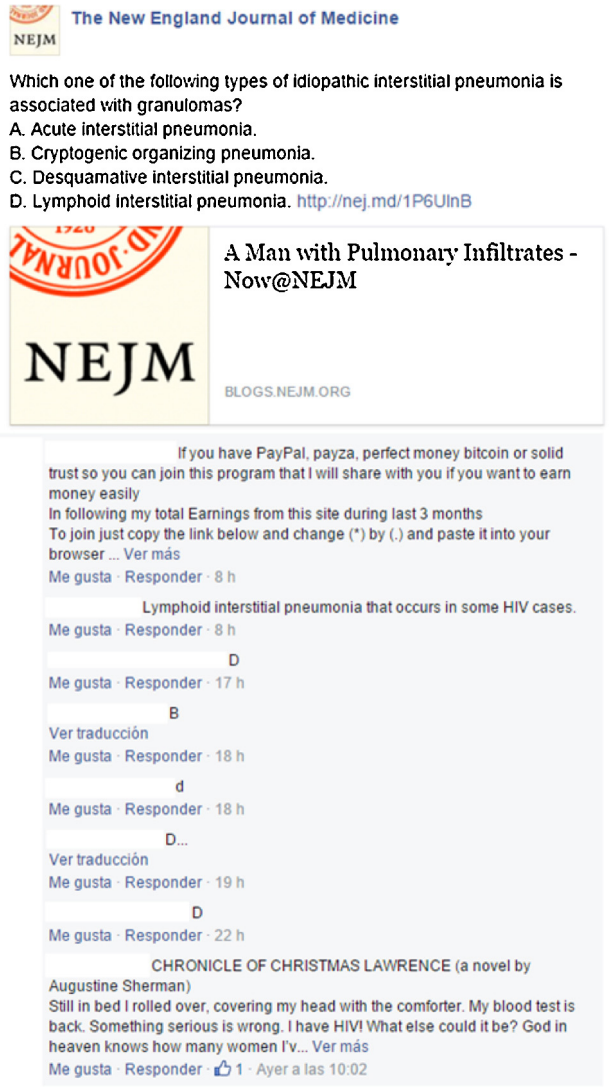


Fig. 1 – Screenshot of a NEJM quiz example.

Once the question is published, users can contribute through their comments. Most of the users try to answer the quiz instead of starting a discussion about the question as the comment analysis confirmed. After a few days, NEJM posts in its timeline the answer to the quiz in a new post.

The benefit behind the effort of NEJM maintaining these quizzes is twofold. On the one hand, it improves the medical education of those users participating and interacting trying to find a correct answer and reading the complementary resources. On the other hand, the comments provided by the different users can be analyzed and mined to extract knowledge hidden in the collective opinion of the users (wisdom of the crowd). The extraction of opinions, feelings or interests is a frequent topic of research in several areas [17–19], used as a basis for developing projects, tools and other efforts oriented to the automatic data extraction from social media [20–23].

Also, the fact that most medical exams around the world are based on medical quizzes increases the interest of the analysis of these kinds of social media platforms quizzes based on intelligent methods.

Please cite this article in press as: A. Rodríguez-González, et al., Automatic extraction and identification of users' responses in Facebook medical quizzes, Comput. Methods Programs Biomed. (2015), http://dx.doi.org/10.1016/j.cmpb.2015.12.025