Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2019-03-20

Deposited version:
Post-print

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Santos, C. L., Rita, P. & Guerreiro, J. (2018). Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. International Journal of Educational Management. 32 (3), 431-447

Further information on publisher's website:
10.1108/IJEM-01-2017-0027

Publisher's copyright statement:
This is the peer reviewed version of the following article: Santos, C. L., Rita, P. & Guerreiro, J. (2018). Improving international attractiveness of higher education institutions based on text mining and sentiment analysis. International Journal of Educational Management. 32 (3), 431-447, which has been published in final form at https://dx.doi.org/10.1108/IJEM-01-2017-0027. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Improving International Attractiveness of Higher Education Institutions based on Text Mining and Sentiment Analysis

Abstract

Purpose – The increasing competition among higher education institutions (HEI) has led students to conduct a more in-depth analysis to choose where to study abroad. Since students are usually unable to visit each HEIs before making their decision, they are strongly influenced by what is written by former international students (IS) on the Internet. HEIs also benefit from such information online. This paper aims to provide an understanding of the drivers of HEIs success online.

Methodology – Due to the increasing amount of information published online, HEIs have to use automatic techniques to search for patterns instead of analyzing such information manually. The present paper uses text mining and sentiment analysis to study online reviews of IS about their HEIs. The paper studied 1938 reviews from 65 different business schools with AACSB accreditation.

Findings – Results show that HEIs may become more attractive online if they financially support students cost of living, provide courses in English, and promote an international environment.

Research limitations – Despite the use of a major platform with a broad number of reviews from students around the world, other sources focused on other types of HEIs may have been used to reinforce the findings in the current paper.

Value – The study pioneers the use of text mining and sentiment analysis to highlight topics and sentiments mentioned in online reviews by students attending HEIs, clarifying how such opinions are correlated with satisfaction. Using such information, HEIs’ managers may focus their efforts on promoting international attractiveness of their institutions.

Keywords International student mobility; Higher education; Text mining; Sentiment analysis; Topic modeling

Paper type Research paper

1. Introduction

Throughout the last decades the promotion of international student mobility (ISM) has been one of the top priorities of the European Commission. ISM strongly benefits both academic institutions and host countries (The European Commission, 2013; Beine, Noël and Ragot, 2014). Students who go through an international experience often remain in the host country after their studies, which contributes to the escalation of highly qualified workers who boost the country's growth (González, Mesanza and Mariel, 2011; Beine, Noël and Ragot, 2014). Furthermore, providing studies to foreign students is a powerful way to communicate the host countries' cultural and political values (González, Mesanza and Mariel, 2011; Beine, Noël and Ragot, 2014). However, Middle East, Asia and Latin America's competition in this field is becoming increasingly more aggressive which leads to a rapid modification of ISM's flows (The European Commission, 2013). To succeed in this competitive environment, Europe should increase HEIs attractiveness even further by effectively promoting ISM. To encourage ISM to their country, higher education institutions (HEI) should acquire a deep knowledge about students' behaviour abroad, exploring the drivers that influence students' decision-making when it comes to choosing their school (Cubillo, Sánchez and Cerviño, 2006; Choudaha and Chang, 2012).

The amount of information about HEIs offering is the main factor that influences the choice of a HEI (Bourke, 2000), especially information posted online because students are unable to visit all the institutions to get to know them in detail (Gomes and Murphy, 2003). International students (IS) often search on independent online sources to obtain information written by former students as they believe it is more trustworthy and detailed than the one provided by HEIs (Gomes and Murphy, 2003). In fact, the Internet is filled with opinions freely expressed by users about their experiences (Mostafa, 2013). Yet, sharing experiences online may strongly influence HEIs reputation which may affect its chances of being selected (Herbig and Milewicz, 1993; Kotler and Fox, 1995; Bourke, 2000; Chun, 2005; Walsh *et al.*, 2009). Freely expressed opinions written online are “unlikely to be biased” (Mostafa, 2013, p. 4241). Thus, the knowledge of IS’ opinions allows institutions to have a more complete vision about the quality of services they provide and also to address some aspects that may be important for their target audience (Feldman, 2013).

The present paper analyses *online* reviews in order to identify the multiple factors that influence the choice of a HEI according to IS’ opinions. Text mining and sentiment analysis techniques were used to deal more effectively with large amounts of data and understand the sentiments of the expressed opinions. Some studies in the literature have used students’ online-expressed opinions (e.g. Isik, Öztaysi, and Fenerci [2012]; Wen, Yang, and Rosé [2014]), but none of them had the goal of understanding HEI’s attractiveness. Such gap in the literature suggests that HEIs may be oblivious to how their reputations are being affected online which may lead to a deterioration of their international appeal.

The present paper contributes to the literature by uncovering the major differences between the different topics mentioned in online *reviews* and also by studying how the students’ sentiments in reviews are correlated with their satisfaction with the HEI they attended. Using such information, the paper may help HEIs’ managers to focus their efforts on promoting international attractiveness of their institutions.

2. Background and theoretical framework

Maringe and Carter (2007, 463) define the process of ISM decision-making as a “multistage and complex process undertaken consciously and sometimes subconsciously by a student intending to enter higher education and by which the problem of choosing a study destination and program is resolved”.

Five major factors able to influence the IS' choice of the destination were determined according to literature: the image of the host country (Bourke, 2000; Mazzarol and Soutar, 2002; Cubillo, Sánchez and Cerviño, 2006); the image of the HEI (Bourke, 2000; Cubillo, Sánchez and Cerviño, 2006); financial factors (Mazzarol and Soutar, 2002; Maringe and Carter, 2007); personal motivations (Bourke, 2000; Maringe and Carter, 2007); and lastly, the amount of information collected is considered to be a fifth factor that crosses the previous ones (Kotler and Fox, 1995; Bourke, 2000).

2.1. Host Country image

Once the student made the decision of studying abroad, the choice of the place of study depends firstly on the country and secondly on the HEI (Bourke, 2000; Mazzarol and Soutar, 2002). Therefore, the host country image plays a vital role in the student decision.

Srikatanyoo and Gnoth (2002, 140) define country's image as the "students cognitive beliefs about the country's degree of industrialisation, national quality standard and other information that is associated with its products and services" meaning that when students choose a country to study, they are not only choosing the education service but also all the other characteristics and services provided by the country. Reputation is one of the elements that influences the country's image formation (Bourke, 2000; Mazzarol and Soutar, 2002; Beine, Noël and Ragot, 2014). According to Bourke (2000), academic education works as a mirror of each country's culture, hence students tend to believe that countries with good reputations provide higher quality education services.

Besides the country, also the characteristics for which the city is internationally recognized may influence the choice since it represents the place where the education service will be provided (Cubillo, Sánchez and Cerviño, 2006).

Other criteria are related to security and discrimination (Mazzarol and Soutar, 2002; Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007), cultural and language distance between host and home countries (Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007; Beine, Noël and Ragot, 2014); the ease of the migration process (Mazzarol and Soutar, 2002; Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007); and the perception of a strong international environment in the host country (Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007; Beine, Noël and Ragot, 2014).

2.2. Institution Image

Srikatanyoo and Gnoth (2002, 140) describe the image of the institution as "students' overall perceptions of institution quality". However, the image of an institution might not correspond to reality, which may cause students who have a negative perception of an institution to avoid it or discredit it (Kotler and Fox, 1995).

A good academic reputation is considered to be one of the main criteria to influence the choice of the HEI and has a very powerful impact on the impression that the student forms about it (Bourke, 2000; Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007; Hemsley-Brown, 2012). To evaluate the academic reputation, rankings have a quite significant influence (Beine, Noël and Ragot, 2014).

The degree's international recognition and its quality are also relevant since the students believe that their chosen degree will eventually boost their opportunities and career progression in the labour market (Bourke, 2000; Mazzarol and Soutar, 2002; Counsell, 2011). The literature also points to teaching quality and experience as important drivers for students to choose a HEI (Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007).

Another essential criteria are: the availability of accommodation (Maringe and Carter, 2007); simply entry procedures (Bourke, 2000); the availability of libraries, study areas

(Bourke, 2000; Price *et al.*, 2003; Maringe and Carter, 2007), technological instruments (Price *et al.*, 2003); and the social life in the institution coupled with the social life of the surrounding city (Price *et al.*, 2003).

2.3.Financial factors

Binsardi and Ekwulugo (2003) showed that the best way to attract IS is to reduce tuition fees or to raise the amount of money spent in international scholarships. However, Beine et al. (2014) show that sometimes students use tuition fees as proxies of institution quality. The higher the tuition fee, the higher the perception of HEI quality. Therefore, students are sometimes willing to pay a more expensive fee in exchange for a top academic education. Besides, the authors also argue that students often do not benefit from scholarships to support their personal expenses, which makes the cost of living an extremely important factor to consider when moving abroad. In this context, commute costs is also considered to be a relevant criteria (Mazzarol and Soutar, 2002; Beine, Noël and Ragot, 2014).

Students not always can afford the tuition fees of a full program; therefore, they need to be sure that they can find work in the host country (Maringe and Carter, 2007). Various countries allow IS to work part-time using their student visas, which is sometimes the only way for them to enjoy an international experience (Mazzarol and Soutar, 2002).

2.4.Personal motivations

Counsell (2011) and Bourke (2000) show that the main reason that leads students to decide to go through an international experience is the belief that they will have better career opportunities in the future. Other students want to study in a specific country simply because they want to go through an international experience (Maringe and Carter, 2007; Counsell, 2011). The desire of learning a second language is often a strong motivation for studying abroad (Bourke, 2000; Counsell, 2011). In fact, students from several countries choose to study overseas so that they can be taught in English (Bourke, 2000).

Other motivations discussed in the literature, are the desire of becoming independent (Bourke, 2000; Counsell, 2011), developing international contact networks (Cubillo, Sánchez and Cerviño, 2006; Maringe and Carter, 2007), and getting to know a different culture (Bourke, 2000; Maringe and Carter, 2007; Counsell, 2011).

2.5.Information

The information obtained about the country, institution and degree shapes the image created by the student throughout the process of selection and has an impact on the student’s decision regardless of its truthfulness (Kotler and Fox, 1995). Kotler and Fox (1995) even claim that information is the foundation of image formation. Such evidence suggests that information is intrinsic to all the other factors mentioned in the literature.

Bourke (2000, 131) argues that “the relevant information which students need to help them decide on a foreign study destination must be communicated using the most effective means possible, taking into consideration language variations and cultural differences.”. However, although the decision of where to study can be affected by the information obtained from institutional sources, recommendations from former students, friends and social networks has a powerful impact on choice (Mazzarol and Soutar, 2002; Maringe, 2006).

Binsardi and Ekwulugo (2003) concluded that most students consider that the best way to promote the institution in order to attract more IS is through the student networks. Indeed, Alumni networks have proven to be one of the most important sources of information to help students’ decide where to enrol in a HEI (Bourke, 2000; Mazzarol and Soutar, 2002; Cubillo, Sánchez and Cerviño, 2006; Wilkins and Huisman, 2014).

Table 1 shows a summary of the factors identified during the literature review.

[Table 1 near here]

Although the present paper argues that all of the above factors are correlated with the satisfaction of students towards a HEI, the different underlying motivations in each factor suggest that there might be significant differences between degrees of satisfaction reported online when students address specific issues in each factor. Therefore, it is proposed that:

H1. The satisfaction of the students towards a HEI differs significantly depending on the factor being addressed.

3. Sentiment analysis on Education and Text Mining Process

In recent years, opinions posted online have been studied using sentiment analysis (SA), a text mining (TM) set of methods to uncover the valence of written text (Mostafa, 2013). TM focuses on automated knowledge discovery from non-structured or semi-structured data collections (Hearst, 1999; Sánchez *et al.*, 2008). Written text is often filled with spelling mistakes, abbreviations and implied meanings. Although humans have the ability to overcome such obstacles, the amount of information online today is so large that it becomes a daunting task to analyse it efficiently without some sort of an automatic pattern discovery technique (Lee *et al.*, 2010; Mostafa, 2013). TM uses natural language processing (NLP) techniques to deal with such challenges and focuses on detecting data patterns automatically (Mostafa, 2013), while SA uses such patterns to identify sentiment polarity in text, for instance, if sentiments are positive, negative or neutral (Mostafa, 2013).

The vast majority of text analysis research in education has been framed within Educational Data Mining (EDM) (Minami and Ohura, 2013). EDM aims to extract new knowledge to improve study environments and students' learning processes (Romero and Ventura, 2013). In recent research, Leong, Lee, and Mak (2012) created a platform that receives messages with students' opinions on lectures, seminars and conferences, and developed an automated method of analysing students' opinions and optimising feedback delivery. Isik, Öztaysi, and Fenerci (2012) used social media reviews and applied TM techniques to understand the most mentioned topics about Istanbul's Technical University. The authors uncovered three major groups of interest, namely "education quality", "campus" and "corporative reputation", and explored general sentiments within each group. The present paper adds further knowledge to the current literature by studying the most mentioned topics in a large group of universities, analysing how certain topics and terms are correlated with students' satisfaction towards the HEIs.

The rationale behind a TM process it is that each document can be represented by the frequency of its terms (Delen and Crossland, 2008). The text in pure ASCII format goes through a preparation stage which includes *stopword* removal (irrelevant terms such as "the" or "and" are taken out) (Delen and Crossland, 2008; Guerreiro, Rita and Trigueiros, 2016). Html tags, whitespaces, numbering and punctuation are also removed to clean noise in the dataset (Guerreiro, Rita and Trigueiros, 2016). A final preparation step includes converting all the characters to lower case and stemming words in the dataset (Guerreiro, Rita and Trigueiros, 2016). Stemming's purpose is to guarantee that words with the same meaning but different suffixes and prefixes are identified as being the same term (stem). For example, words such as "wonderfully", "wonderful", "wonders" are all associated to the stem "wonder" after the stemming transformation (Porter, 1980). Part-of-Speech (POS) is a technique used occasionally, ideally before stemming and punctuation removal. POS' purpose is to tag a syntactic class to a word so that only some specific grammatical categories are selected for analysis (Jackson and Moulinier, 2007). Data is then converted into a document-by-term matrix (DTM) in which each column represents a term, each line represents a document and each cell has the term's frequency in the document (Delen and Crossland, 2008; Grün and Hornik, 2011). Some terms appear multiple times in several documents, but most terms appear only in a few documents and the matrix can be quite sparse (Guerreiro, Rita and Trigueiros, 2016). In such cases, the term-frequency-inverse document frequency (TF-IDF) should be used instead of the term frequency to highlight the most discriminative terms (Delen and Crossland, 2008; Feinerer, Hornik and Meyer, 2008; Guerreiro, Rita and Trigueiros, 2016). The DTM is the starting point for text mining exploration of patterns in data. After the non-structured text has been converted to a structured matrix, clustering algorithms such as topic models are often used to uncover the correlation between the various terms in the documents.

Topic models are mixed-membership models in which each document belongs to several latent topics with different distributions (Grün and Hornik, 2011). One example of a topic model is the Biterm Topic Model (BTM). BTM extracts all word pairs (biterms) per document and calculates the chances of each biterm belonging to each latent topic (Cheng *et al.*, 2014). Finally, topics are profiled and labelled according to the most correlated terms in each one (Guerreiro, Rita and Trigueiros, 2016). Recent studies show BTM is fit to deal with short textual content such as twitter messages or small reviews (Cheng *et al.*, 2014).

A final step to understand written text is to explore the sentimental valence of the document, sentence, or a specific term (Liu and Zhang, 2012; Feldman, 2013). During sentiment analysis two approaches are used to classify sentiments – machine learning or lexicon-based (Liu and Zhang, 2012; Feldman, 2013). In the first one, traditional mining algorithms are used, such as support vector machines and naive bayes. In the second one, dictionaries of terms such as WordNet (dictionary-based approach) or other lexicons of a specific domain (corpus-based approach) are used to classify the sentiments in text. Some commercial web-services are also available to deal with sentiment analysis. For example, some recent studies use Semantria[1] successfully to deal with sentiment analysis (Gao, Hao, and Fu 2015; Serrano-Guerrero et al. 2015; Peisenieks and Skadiņš 2014). Through an Excel plug-in, Semantria prepares the data and extracts the sentiments of the text corpus. Although Semantria

[1] https://www.lexalytics.com/

is a black-box, it uses approximately 40 different algorithms, including NLP techniques and big word lexicons (Lexalytics, 2015c) to analyse text. Semantria identifies sentiment expressions (such as noun-adjective combinations), calculates the sentiment score of text through a logarithmic scale between -10 and 10, and finally classifies text with a sentiment polarity (positive, negative or neutral) according to the obtained score.

The present paper uses BTM to identify the different topics discussed by online students and Semantria to classify the sentiments expressed by IS about their HEI.

4. Method

4.1.Data selection and extraction

Online opinions from students around the world were collected from the social platform iagora.com[2]. Iagora asks students to write about different aspects of their international experience – housing, student life, expenses, academic, language, overall, and final comments – and then asks students to classify several sub-criteria for each aspect in a quantitative scale. Each aspect has a final score calculated using the average of those sub classifications and a global scale for each aspect that ranges from 1-5.

To control for the possibility that students from different areas of study might have different motivations to choose an HEI, only students from the same area were considered. Thus, only reviews of students who chose to study in a business-school were analysed. The paper collected information about all universities with an Association to Advance Collegiate Schools of Business (AACSB) accreditation until May 2015. The final dataset had 65 European business-schools with student-formed opinions. 1938 reviews were extracted and a manual

[2] http://www.iagora.com/studies/

random verification of 10% of the reviews was made in order to confirm the coherence of the extracted data.

4.2.Data Preparation

The first task in data preparation was to deal with data quality issues. Reviews with random characters and reviews related to student mobility inside their home country were removed. The final dataset ended up having a total of 1925 reviews with a written opinion of students studying abroad.

The reviews were analysed through QDA Miner tool[3] to detect spelling mistakes. Figure 1 presents the replacements suggested by QDA Miner and used in the paper.

[Figure 1 Replacements suggested by QDA Miner]

All htmls and numbering were removed, using the *gsub* package of the statistical software R. Multiple whitespaces and repeated punctuation signs were also replaced by only one. In line with Hu and Liu (2004), POS was used to focus the analysis on nouns aiming to get a better contextualisation of the issues that students are talking about regardless of their sentiments. Terms with /NN (singular nouns) and /NNS (plural nouns) tags were extracted.

Stopwords were removed as well as the rest of the punctuation. All words were converted to lowercase and stemming was applied. Finally, a DTM was created from the student reviews.

DTM was quite sparse (100%) with a total of 1925 documents and 4629 terms. In order to reduce sparsity, only terms which occurred at least twice in all the reviews were kept. In line with Grün and Hornik (2011), a minimum three-character size per term was defined and TF-IDF was applied. Terms in which TF-IDF score was slightly inferior to the median as well as

[3] http://provalisresearch.com/

documents in which the sum TF-IDF scores were below or equal to zero were also removed (Grün and Hornik 2011). After reducing sparsity, the final DTM had 1853 documents with 1799 terms.

Figure 2 shows a wordcloud of stemmed terms created using the DTM and suggests that the reviews are especially focused on the campus (campus), accommodation (accommod) and the school (school). A closer look also highlights terms around expenses (for example, cost; price; money), culture (cultur, travel), academic matters (teacher, class), leisure (pub, cafe, parti) and even information (inform).

[Figure 2 Wordcloud of the most frequent stemmed terms]

4.3.Topic Modelling

The BTM algorithm was used as a topic modelling technique to uncover latent topics in the text. To select the number of k topics to find, perplexity and log-likelihood measures were calculated for a different number of topics. A suitable number of topics is reached when there are no significant changes on the model's explained variability as the topic quantity rises (Guerreiro, Rita, and Trigueiros 2015). According to Figure 3, ten topics were selected for analysis.

[Figure 3 Log-Likelihood and Perplexity]

Table 2 shows the results of a profiling analysis of the ten topics obtained using BTM.

[Table 2 near here]

Table 3 shows the distribution of the most correlated topic with each review. *Student Life* is the topic with most reviews. This happened since this topic is associated to many aspects of the experience very fairly. To better distinguish the reviews through a fairer distribution, we chose to suppress the Student Life topic from the analysis. Thus, the reviews correlated with this topic in the first place were associated to the second most correlated one, and the reviews which were not associated with it in the first place remained with the most correlated topic according to BTM. Table 4 shows the topic distribution used in the following analysis.

[Table 3 near here] **[Table 4 near here]**

4.4.Topic Profiling

The topics' assignments were based on the most correlated terms and after analysing the most correlated reviews with each topic.

Housing: In this topic the two words are related to accommodation, namely: *apart* from the word apartment and *resid* from the word residence (usually associated with student residences). The stem *accommod* (accommodation) is also present in the topic. The literature review suggests that one of the main criteria to choose a HEI is its accommodation or at least the information about it (Maringe and Carter 2007). In online reviews, students talk about the type of accommodation (if it was offered by the campus or if it was an independent accommodation) and also about the facilities: the quality of the surrounding environment, the distance to the campus or to the city centre (stems *town* e *centr*), and the international environment (*campus* e *exchang* stems).

Campus facilities: In this topic, the first six terms all discuss facilities, such as the kitchen, bathroom and toilet or even sport facilities, suggesting the importance of facilities to students. The literature highlights the importance of facilities such as libraries and study areas (Bourke 2000; Maringe and Carter 2007; Price et al. 2003). According to the results, the most frequently reviewed facilities are the ones from the campus' accommodation. We also encountered many reviews discussing sport facilities.

Culture enrichment: The most correlated term with this topic is the *cultur* stem, which suggests that students believe international mobility is an opportunity (*opportune* stem) to get to know a new culture. The *host* stem appears usually related to the “host country” or to “host family”. This topic is focused on reviews that emphasize the opportunity of enjoying the international experience on a multicultural level. In the literature, the desire of meeting new cultures is pointed out as one of the main motivations for an international experience (Bourke 2000; Counsell 2011; Maringe and Carter 2007). Meeting new people is related to the desire to broaden the international contacts network (Cubillo, Sánchez, and Cerviño 2006; Maringe and Carter 2007).

Host city life: This topic reveals the interest of students about the host city’s life style, such as opinions about pubs (*pub* stem), bar’s (*bar* stem), cafés (*cafe* stem) and parties (*parti* stem) in the town centre (*town centr* stems). The literature shows that a city´s reputation as a world renowned place could attract students (Cubillo, Sánchez, and Cerviño 2006). However, according to our results, the reviews seem to refer to the city in a more festive and bohemian way than in a cultural one. We notice the importance given to the distances between the city centre and the campus as well as the accommodation which suggests the interest given to their location in relation to the city centre. The *accent* stem highlights the difficulties some international students have to understand the local population’s accent. According to previous literature, language proximity is one of the important drivers in selecting a HEI, since it makes it easier for students to communicate (Beine, Noël, and Ragot 2014; Cubillo, Sánchez, and Cerviño 2006; Maringe and Carter 2007).

Academic: This topic is clearly related to academic aspects which we can easily identify through stems such as *exam*, *professor*, *class*, *test*, *lectur*, and others. There are reviews related to teaching and other academic matters such as classes and exams. The literature argues that a good academic reputation as well as the quality of the professors are decisive criteria for choosing a HEI (Maringe and Carter 2007; Mazzarol and Soutar 2002).

Expenses: The cost of living is the main subject discussed in the reviews associated with this topic. Students discuss about commute costs both inside the host city and to other foreign countries. For example, they share tips on how to save money through discounts (*discount* stem) or the purchase of train cards (*card, train*, *travel* stems) and tips on how to manage money through bank accounts (*bank* and *account* stems). However, as expected, no reviews talked about academic expenses. In fact, Beine et al. (2014) show that the cost of living is more important in choosing a HEI than academic expenses.

Tourism: Many reviews on this topic encouraged students to take advantage of the host country's geographic location to travel and get to know other countries. Therefore, we believe that the geographic location of the country might also influence the choice of the destiny of the HEI. For example, for having boarders with several countries, Switzerland might be considered more attractive than Portugal, since the latter boarders with only one country.

Skills development: Skills development were frequently mentioned in the *reviews*, especially the motivation to master a new language. *Level* stem often appears in reviews in which students confirm that they have chosen a certain country to master the country's language. The *studi* stem can be found on reviews who discuss the need to learn the language before travelling to the host country. An example is the review which states: "*before coming to france study french*". The literature review shows that one of the main personal motivations to choose a HEI is to improve linguistic skills (Bourke 2000; Counsell 2011). The English language is the most prevalent language discussed in the topic since several reviews talk about the academic programs being taught in English. The *program* stem shows up correlated to language aspects like in the following review: "*The program is in English, but it is better to learn some French*". Apart from language discussions, many reviews also mention the quality of the programs and of the institutions' teaching.

International environment: This topic underlines the importance of the international environment (both academic and social environment). On the academic environment discussions, students seem particularly worried about the language in which courses are taught. They prefer classes to be taught in English. Although the current paper only studied opinions wrote in English, which could lead students to address that particular interest in their discussions, findings are aligned with the literature that shows that the desire to learn English as a second language is a strong motivation for studying abroad (Bourke, 2000; Counsell, 2011).

On the social environment, students mainly discuss the diversity of social activities for international students, such as parties. Many authors have already referred the importance of the promotion of an international environment in order to attract IS (Beine, Noël, and Ragot 2014; Cubillo, Sánchez, and Cerviño 2006; Maringe and Carter 2007).

Table 5 presents a matching between what literature has found to be more important when choosing an HEI and what students write on the online reviews, as discussed above.

[Table 5 near here]

5. Results

5.1.Hypothesis Testing

To validate **H1,** the overall satisfaction degree of students regarding their HEI, ranging from 1 to 5 was coded as the dependent variable (OV_STARS). Figure 4 shows the dependent variable distribution.

[Figure 4 Ov_stars distribution]

Kruskal-Wallis was conducted to test if there were any significant differences in overall satisfaction between the topics. Kruskal-Wallis test was used since the dependent variable is not continuous and does not follow a normal distribution. The non-parametric Levene test (Nordstokke and Zumbo 2010; Nordstokke et al. 2011) confirmed the homoscedasticity of variances with p = 0.984.

Table 6 shows that *Culture Enrichment* and *Academic* topics have the highest and the lowest mean rank respectively and that there is a significant difference in satisfaction between groups (H(8)=20.681, p=.008), thus supporting **H1**.

[Table 6 near here]

The Dunn-Bonferroni post-hoc test was used to compare groups in pairs. There is only a significant difference between the *Housing* (mean rank=843.92) and the *Culture Enrichment* (mean rank=973.53) topics with p=0.028[4]. Results show that students that discuss cultural enrichment in their *reviews* tend to feel more satisfied with their international experience than those who approach subjects related to accommodation. There were no significant differences in satisfaction on the remaining topics.

5.2.Sentiment Analysis

Semantria detects the sentiment of many instances such as: entities, entity types, themes (Lexalytics 2015a, 2015b, 2015c). According to Lexalytics (2015a), entities refer to proper nouns, such as a person (“Steve Jobs”), a place (“Denmark”), a product (“iPhone”). Each entity is associated to an entity type; for instance, the “Denmark” entity is related to the “Place” entity type. Themes refer to other nouns representing more generic concepts people talk about, for example “international student”.

Figure 5 shows a wordcloud presented by Semantria in which IS appears as the main term. Most terms in the wordcloud fit in the BTM topics. There are many references to expenses

[4] See Appendix 1

(really expensive; quite expensive; bank account), cultural enrichment (different culture, local people), accommodation (student house, private accommodation), skills development (language skills, communication skills), academic matters (academic life, great university), city life (city center, night spots, night life), tourism (different countries) and international environment (international students). Sentiments and their strength are shown using a colourful spectrum that ranges from red (negative sentiment) to grey (neutral sentiment) and green (positive sentiment).

Results suggest that students' discontent is mainly due to perceived difficulties and expenses. Term bank account, however, is positive going along to the results obtained within the topics when students advise to open bank accounts as a better way to manage their money. Moreover, students' satisfaction shows up mainly related to the international and multicultural environment as well as social life in the city.

[Figure 5 Semantria wordcloud]

6. Conclusion

International students often rely on online sources to select a HEI abroad since they are unable to visit the school and because they find ex-students' opinions to be trustworthy (Gomes and Murphy, 2003). These sources of information may strongly influence the reputation of a HEI and consequently the decision-making of students. Thus, it becomes essential for an institution to be aware of students' opinions on social media so that they can manage their reputation and meet IS needs and wants.

However, the amount of information published online is so vast that analyzing it manually becomes an extremely time consuming process.. Using text mining techniques and sentiment analysis, the present paper focused on analysing the most prominent topics discussed in online reviews written by international students to help HEI's managers understand the real issues being discussed IS.

As academic contributions, the current paper supported the hypothesis H1 that the satisfaction of the students towards a HEI is significantly different depending on the factor being discussed in their online opinion.

Findings show that students are significantly more satisfied about their international experience when discussing topics around cultural enrichment than when discussing accommodation issues. Findings agree with the literature that shows that the desire to meet new cultures is one of the main motivations for taking an international experience (Bourke 2000; Counsell 2011; Maringe and Carter 2007). While the first topic is a strong motivation for going abroad according to the literature, the second relates to a more functional side of staying at a foreign country.

Concerning accommodation issues, the availability of the accommodation by the HEI was considerably emphasized, as proposed by Maringe and Carter (2007). Reviews suggested that students have a special interest in staying in accommodations provided by the campus since they want to cohabit with a larger quantity of multicultural students.

Results also show that financial aspects and personal motivations are the topics that most influence students' satisfaction towards a HEI. The most outstanding costs are related to the host country's cost of living and commutes as suggested by Beine et al. (2014). Results did not reveal major references relative to the cost of HEI's tuition fees.

Regarding personal motivations, the importance of skills development especially at a linguistic level as enhanced by Bourke (2000) and Counsell (2011) was also highlighted. We verified, in this case, that the English language has a primordial place. Evidences also revealed the desire of being integrated into an international environment while enhancing the contacts network. The contact with people from different cultures was one of the most stressed elements on the outcomes.

That being said, as practical contributions we recommend that HEIs offer a higher financial support when it comes to living costs if they want to be considered more attractive. It would also be beneficial to support students who want to find a part-time job.

Academically, it is indispensable to make available courses in English. It is also advantageous that the HEI provide accommodation on campus. If it is not possible, then they must offer help to find housing. Promoting social activities for IS would also be very appealing.

Apart from the efforts made by the HEI to provide what was recommended, it is imperative that the students easily perceive it. Thus, HEI must clearly make this information available and invest in marketing to enhance it, as suggested by Bourke (2000). Moreover, it is fundamental that marketing and communication include both the academic and social side, including the promotion of an international environment and characteristics of the country as also suggested by Cubillo et al. (2006).

A limitation of the current study is focused on the quality of the information reported on recommendation platforms. Although this kind of platforms often require students to have their own login and password in order to leave their opinion, they are not free from the same plagues that often bias recommendation sites, such as false registrations and testimonials to increase institutions ratings.

For future research we consider to be interesting to better explore if the existence of sport areas, the location of the accommodation and of the campus in relation to the city centres, as well as the geographical location of countries may be relevant criteria when choosing a HEI, as the results suggested.

Bibliography

Beine, M., Noël, R. and Ragot, L. (2014) 'Determinants of the international mobility of students', *Economics of Education Review*, 41(2014), pp. 40–54.

Binsardi, A. and Ekwulugo, F. (2003) 'International marketing of British education: research on the students' perception and the UK market penetration', *Marketing Intelligence & Planning*, 21(5), pp. 318–327.

Bourke, A. (2000) 'A Model of the Determinants of International Trade in Higher Education', *The Service Industries Journal*, 20(1), pp. 110–138.

Cheng, X., Yan, X., Lan, Y. and Guo, J. (2014) 'BTM: Topic Modeling over Short Texts', *IEEE Transactions on Knowledge and Data Engineering*, 26(12), pp. 2928–2941.

Choudaha, R. and Chang, L. (2012) *Trends in international student mobility, World Education News & Reviews*. World Education Services, New York.

Chun, R. (2005) 'Corporate reputation: Meaning and measurement', *International Journal of Management Reviews*, 7(2), pp. 91–109.

Counsell, D. (2011) 'Chinese Students Abroad: Why They Choose the UK and How They See Their Future', *China: An International Journal*, 9, pp. 48–71.

Cubillo, J. M., Sánchez, J. and Cerviño, J. (2006) 'International students' decision-making process', *International Journal of Educational Management*, 20(2), pp. 101–115.

Delen, D. and Crossland, M. D. (2008) 'Seeding the survey and analysis of research literature with text mining', *Expert Systems with Applications*, 34(3), pp. 1707–1720.

Feinerer, I., Hornik, K. and Meyer, D. (2008) 'Text Mining Infrastructure in R', *Journal of Statistical Software*, 25(5), pp. 1–54.

Feldman, R. (2013) 'Techniques and applications for sentiment analysis', *Communications of the ACM*, 56(4), pp. 82–89.

Gao, S., Hao, J. and Fu, Y. (2015) 'The Application and Comparison of Web Services for Sentiment Analysis in Tourism', in *12th International Conference on Service Systems and Service Management (ICSSSM)*. Guangzhou, China: IEEE, pp. 1–6.

Gomes, L. and Murphy, J. (2003) 'An exploratory study of marketing international education online', *International Journal of Educational Management and Marketing*, 17(3), pp. 116–125.

González, C. R., Mesanza, R. B. and Mariel, P. (2011) 'The determinants of international student mobility flows: An empirical study on the Erasmus programme', *Higher Education*, 62, pp. 413–430.

Grün, B. and Hornik, K. (2011) ‘Topicmodels : An R Package for Fitting Topic Models’, *Journal of Statistical Software*, 40(13), pp. 1–30.

Guerreiro, J., Rita, P. and Trigueiros, D. (2016) ‘A Text Mining-Based Review of Cause-Related Marketing Literature’, *Journal of Business Ethics*, 139(1), pp. 111–128.

Hearst, M. (1999) ‘Untangling text data mining’, in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. University of Maryland, USA: Association for Computational Linguistics, pp. 3–10.

Hemsley-Brown, J. (2012) ‘The Best Education in the World: reality, repetition or cliché? International students’ reasons for choosing an English university’, *Studies in Higher Education*, 37(8), pp. 37–41.

Herbig, P. and Milewicz, J. (1993) ‘The relationship of reputation and credibility to brand success’, *Journal of Consumer Marketing*, 10(3), pp. 18–24.

Hu, M. and Liu, B. (2004) ‘Mining opinion features in customer reviews’, in *Proceedings of the National Conference on Artificial Intelligence*. San Jose, California: AAAI Press, pp. 755–760.

Isik, M., Öztaysi, B. and Fenerci, K. H. (2012) ‘A Sentiment Analysis as a Tool to Identify The Status Of Universities: The Case of ITU’, in *Procedings of the 2012 International Conference on Industrial Engineering and Operations Management*. July 3 – 6, Istanbul, Turkey, pp. 1118–1126.

Jackson, P. and Moulinier, I. (2007) *Natural Language Processing for Online Applications: Text retrieval, extraction and categorization*. 2nd edn. Amsterdam: John Benjamins.

Kotler, P. and Fox, K. (1995) *Strategic Marketing for Educational Institutions*. 2nd edn. Englewood Cliffs, NJ: Prentice-Hall.

Lee, S., Baker, J., Song, J. and Wetherbe, J. C. (2010) ‘An Empirical Comparison of Four Text Mining Methods’, in *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS)*. IEEE, pp. 1–10.

Leong, C. K., Lee, Y. H. and Mak, W. K. (2012) ‘Mining sentiments in SMS texts for teaching evaluation’, *Expert Systems with Applications*. Elsevier Ltd, 39(3), pp. 2584–2589.

Lexalytics (2015a) *Categorization of Text*. Boston.

Lexalytics (2015b) *Entity Extraction*. Boston.

Lexalytics (2015c) *What is Sentiment Scoring?* Boston.

Liu, B. and Zhang, L. (2012) ‘A survey of opinion mining and sentiment analysis’, in Aggarwal, C. C. and Zhai, C. X. (eds) *Mining Text Data*. New York: Springer US, pp. 415–460.

Maringe, F. (2006) ‘University and course choice: implications for positioning, recruitment and marketing’, *International Journal of Educational Management,* 20(6), pp. 466–479.

Maringe, F. and Carter, S. (2007) ‘International students’ motivations for studying in UK HE: insights into the choice and decision making of African students’, *International Journal of Educational Management*, 21(6), pp. 459–475.

Mazzarol, T. and Soutar, G. N. (2002) ‘“Push-pull” factors influencing international student destination choice’, *International Journal of Educational Management,* 16(2), pp. 82–90.

Minami, T. and Ohura, Y. (2013) ‘Lecture Data Analysis towards to Know How the Students’ Attitudes Affect to their Evaluations’, in *The 8th International Conference on Information Technology and Application*. Sidney, Australia, pp. 164–169.

Mostafa, M. M. (2013) ‘More than words: Social networks’ text mining for consumer brand sentiments’, *Expert Systems with Applications*, 40(10), pp. 4241–4251.

Nordstokke, D. W. and Zumbo, B. D. (2010) ‘A new nonparametric levene test for equal variances’, *Psicologica*, 31(2), pp. 401–403.

Nordstokke, D. W., Zumbo, B. D., Cairns, S. L. and Saklofske, D. H. (2011) ‘The operating characteristics of the nonparametric Levene test for equal variances with assessment and evaluation data.’, *Practical Assessment, Research & Evaluation*, 16(5), pp. 1–8.

Peisenieks, J. and Skadiņš, R. (2014) ‘Uses of Machine Translation in the Sentiment Analysis of Tweets’, in *Human Language Technologies-The Baltic Perspective: Proceedings of the Sixth International Conference Baltic HLT*. Kaunas, Lithuania: IOS Press, pp. 126–131. d

Porter, M. F. (1980) ‘An algorithm for suffix stripping’, *Program: electronic library and information systems,* 14(3), pp. 130–137.

Price, I., Matzdorf, F., Smith, L. and Agahi, H. (2003) ‘The impact of facilities on student choice of university’, *Facilities,* 21(10), pp. 212–222.

Romero, C. and Ventura, S. (2013) ‘Data mining in education’, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 3(1), pp. 12–27.

Sánchez, D., Martín-Bautista, M. J., Blanco, I. and Torre, C. J. D. La (2008) ‘Text Knowledge Mining: An Alternative to Text Data Mining’, in *2008 IEEE International Conference on Data Mining Workshops*. Pisa, Italy: IEEE, pp. 664–672.

Serrano-Guerrero, J., Olivas, J. A., Romero, F. P. and Herrera-Viedma, E. (2015) ‘Sentiment analysis: A review and comparative analysis of web services’, *Information Sciences,* 311, pp. 18–38.

Srikatanyoo, N. and Gnoth, J. (2002) ‘Country image and international tertiary education’, *Journal of Brand Management,* 10, pp. 139–146.

The European Commission (2013) *O Ensino Superior Europeu no Mundo, Comunicação da Comissão ao Parlamento Europeu, ao conselho, ao Comité Económico e Social Europeu e ao Comité das Regiões*. Vol. COM(2013) 499 Final. Brussels.

Walsh, G., Mitchell, V.-W., Jackson, P. R. and Beatty, S. E. (2009) ‘Examining the Antecedents and Consequences of Corporate Reputation: A Costumer Perspective’, *British Journal of Management,* 20(2), pp. 187–203.

Wen, M., Yang, D. and Rosé, C. (2014) ‘Sentiment Analysis in MOOC Discussion Forums: What does it tell us?’, in *Proceedings of the 7th International Conference on Educational Data Mining*. London, UK, pp. 130–137.

Wilkins, S. and Huisman, J. (2014) ‘Factors affecting university image formation among prospective higher education students: the case of international branch campuses’, *Studies in Higher Education,* (December), pp. 1–17.

Appendices

Appendix 1

[Table 7 near here]