

The Use of Social Media and Internet Data-Mining for the Tourist Industry

Krsak B* and Kysela K

Faculty of Mining, Ecology, Process Control and Geotechnology, Technical University of Kosice, Slovakia

Abstract

Increasing competition in the field of the provision of services in tourism is forcing companies and institutions providing these services to think about the possibilities to increase their competitiveness, efficiency, and productivity. The aim of this initiative is to increase the customer's satisfaction with the services provided, which directly correlates with the increase in the market share and viability of the organization. One of the important ways of enhancing the competitiveness is the collection, analysis, and application of data of the target groups. The source of these data is mostly Internet and the web-based services and solutions, such as Google Trends, Google search engine, Google Analytics, Flickr, social networks, etc. Therefore, we have conducted a comparative analysis by studying primary and secondary sources to find out which source of information provides the most interesting data to be mined in the area of the tourist industry by analyzing the searchability of the tourist destination through a set of pre-defined keywords. Further to a collection of the sources and their analysis, we have compared them by counting the amount of touristsearched relevant keywords and developed conclusions. Our study shows that the greatest impact the searchability of a tourist destination has on the tourist location is virtual communities and reviews found on the Internet.

Keywords: Depth data analysis; Data-mining; Social networks; Google analytics; Tourist industry; Google trends

Application of Data-Mining in Tourism

All operators in the tourism industry, including governmental and non-profit organizations, providing cultural and tourist information, accommodation and other related services need to be able to define relations between tourism activities and preferences of tourists in order to plan the required infrastructure, such as transport and accommodation. They also need a detailed analysis of operational and strategic planning decisions in the future. An example would be investments in projects, distribution of resources such as human resources and time, as well as marketing plans and web portals, brochures and other marketing tools. To meet these objectives, it is imperative to apply the output of data-mining. According to Bose and Mahapatra from Hong Kong University, Data-mining may be used in three key areas of tourism, namely:

- (1) Forecasting expenses on tourism
- (2) Analysis of the tourist profile as a target group and
- (3) Forecasting the number of tourist arrivals.

The given methods and procedures of use of data-mining and information technology are described hereunder [1].

The Definition of a Data-Mining Process

Data-mining is a demanding process, which is used in various fields. Is it possible to meet with this term in both, finances and banking, but also in telecommunications, the biopharmaceutical industry, security technology and other sectors where it is necessary to efficiently process and mine data for later use in the decision-making process [1,2]. As the technology is moving quickly and the data volume thus also increases, it is creating a need for a complex of methods and models, such as the amount of traffic for a reasonable time to get the relevant results based on the input conditions.

Data-mining involves four basic steps, namely: (1) data collection, (2) data cleaning (3) data analysis and (4) interpretation and evaluation [1,2]. Data collection process requires extracting data from the most relevant sources. The expansion of the Internet and information technologies provide for many relevant sources of data, such as social networks, frequency of keyword search using search engines as a marketing tool e.g. Google Analytics, Google Ad Words. Data cleansing process is ensuring that the collected data are consistent and properly recorded [3]. In this step, it is detected by common data errors, which are subsequently replaced or repaired. Thirdly, and most important step is a subsequent interpretation of the data using statistical and information methods and techniques. The choice of these methods depends on the type of problem and also the availability of appropriate software for the data-mining analysis [1,4]. The most difficult step is to analyze the outcomes of interpreted data - which may indicate a high degree of accuracy, however, may not be in direct correlation to the issue, which was data-mined. When the results of this procedure indicate a meaningful interpretation, the next step is the application in practice (Figure 1).

The Concept of Data Mining

Regarding the concept of data mining, the definition states that data mining is the process of analyzing data from different perspectives and their conversion into useful information. From a mathematical and statistical point of view, it is about finding correlations, namely the interrelationships or patterns in the data. The information obtained is then used in decision-making and their use should be measured in the achieved economic effect. Data-mining also can help identify the problem and identify existing or likely interrelationships between the Entities [1,2,5].

*Corresponding author: Krsak B, MSc., PhD., Faculty of Mining, Ecology, Process Control and Geotechnology, Technical University of Kosice, Letna 9, 042 00 Kosice, Slovakia, Tel:+421 55 602 2664; E-mail: branislav.krsak@tuke.sk

Received December 22, 2015; Accepted February 17, 2016; Published February 25, 2016

Citation: Krsak B, Kysela K (2016) The Use of Social Media and Internet Data-Mining for the Tourist Industry. J Tourism Hospit 5: 197. doi:10.4172/2167-0269.1000197

Copyright: © 2016 Krsak B, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



Use of Data Mining in Tourism - Social Networks

The term social network could be generally understood in the form of Internet applications, which carry the user content that includes media impressions created by users, usually influenced by relevant experience and experiences that are shared online for their easier access by other users [6]. Social networks thus include a myriad of applications that allow individual users to publish, label, or blog about their impressions, experiences, adventures, but also unconfirmed rumours on the Internet. Immediately generated user content represent a new source of information that circulate through virtual space to educate and inform other users about various services, products, trademarks, or other matters [1]. This data, unlike the information produced by manufacturers and service providers, is made available directly to customers and is shared among communities. This collective information benefits more and more tourists each day, making it increasingly difficult for businesses to market tourism.

Tourist's virtual communities, such as Lonely Planet and IGoUgo, where tourists can exchange views and experiences about common interests, are available on the Internet since the late 90s a number of experts have tried to analyze their impact in connection with the tourism. Nowadays, many online tools are available, such as sharing and social networking sites and various applications with media content (e.g. Trip Advisor), Internet forums, rating and customer feedback, evaluation systems (e.g. In smart app booking. com) Virtual Worlds (e.g. SecondLife), podcasts, blogs and online videos–vlogs [7]. It is through these internet-supported views and shared experiences are collected data and information that are used to enhance and improve services for tourists.

As we can see, for the success of websites such as tripadvisor.com and zagat.com, it is just ratings and reviews of customers and tourists, which often form the basis for future buyer's decisions on buying or not buying. Research of this type of social media and networks point to a high degree of impact and evaluation reviews the tourists can decide about [8-10].

Sharing Pictures Potential

An important step leading to the promotion of tourism is also an effort to support and understand the behaviour of tourists and analyze

the exact address and the correct choice of destination. The capture of information flow and the movement of tourists are possible using so-called Website Hosting that allow users to upload their photos and locate each of them on the map. The main idea is to mark (check-in) at the specific locations where the user is located [11]. Among the major websites that allow such service, we include Flicker and Panoramio. Users of these domains add photos and geographical attributes. Google Maps is used to pinpoint the user's location. Na photos in addition to the physical attributes and displays the time it was at the photo added, it is also possible to analyze the area by year, according to the season, time of the day, etc. . One of the main expectation is to separate tourists by country and the city, where they come from, these significant data can be selected by using the domain Flickr. To determine the length of stay in the country, the date of adding the first picture to the last picture is used every now and again. Simultaneously, it examines whether, and in what period of time, there is a reintroduction photo of the same place and actually can find out if a tourist has repeatedly visited and for how long the same place [11].

Use of Data-Mining in Tourism - Google Keywords (keywords)

Research on the use of social networking and information technology using data mining tools and procedures by, shows the importance of the interpretation of the results for tourism and tourism. In this study, the authors tried to imitate the planning of a random tourist using an online search engine (Google) when searching for information about destinations [12]. They chose the methodology of data-mining, and data collection, data cleansing and subsequent evaluation and interpretation of the data to convert the selected frequency of keywords entered into the search engine Google. The aim was to examine the usability aspects of social media and networks based on the specified entry requirements. Done aspects include (1) the share of social media in search results from search engine (Google), (2) the way they were represented in social media in search results, and (3) the type of websites with social media, and (4) the relationship between keyword searches and types of websites with social media.

For this analysis was chosen the 10-set of predefined keywords in combination with nine destinations. Key words contained "accommodation", "hotel", "activities", "points", "park", "events ", " tourism ", " restaurant ", " shopping " and " night life ", which are the most common search terms related to tourism used by tourists who seek information relating to tourism in the destinations. The choice of these words was based on earlier studies that were intended to reflect generic keywords and general categories of words related to tourism [5,7]. The research focuses on destinations that appeared in search results as a constant. 9 destinations were selected in the United States, regarding the largest to the smallest, the amount of traffic, reflecting geographic diversity. These destinations were New York, Chicago, LasVegas, Dallas, Charlotte (NC), San Jose (CA), Elkhart (IN), Bradenton (FL), and Pueblo (CO). This selection was considered appropriate to the nature of the research. The city names as keywords were assigned abbreviation of the state to prevent the occurrence of irregularities in the city of the same name in another country [13].

Search engine Google was chosen because it represents one of the most modern and most popular search technologies in the search technologies market. At the given time, Google handles the largest share of search requests (approximately 47. 3%) on the Internet and scans more than 25 billion websites in the context of more than 250 million search queries a day [14,15]. Google is the most dominant in

Page 2 of 3

the United States, where he manages nearly two-thirds of all online search queries. Specifically, in the tourism sector and tourism, Google is among the top 10 Internet portals that generate the most customer

The analysis of the 10-predefined keywords (Figure 2), compared with the 9 destinations were different websites in search results. Internet pages containing thousands of keyword combinations emerged in the results. Authors of this study took into account the first 5 pages in the Google search engine, under the premise that the most users are less likely to browse more than 5 pages in search engines at once. The results show that the largest number of sites with keywords are pages of virtual communities (40%), reviews 27%, blogs (15%), social networks (9%), media sharing 7% and others 2% [17].

Use of Data Mining in Tourism - Google Trends

sites for individual service providers [16].

Google Trends measures the likelihood of searches. Google Insights for Search" analyses a portion of Google web searches to calculate the number of searches that have been entered by the user for specific terms, relative to the total number of searches for the same term on Google over time. This given analysis indicates the likelihood of a user searching for a certain given term from a specific location at a time. Google's system eliminates the same entries from a single user over a short time so that it prevents low-quality outcome [18].

The relative data in "Insights for Search" are displayed on a scale of 0 to 100 and each point has been graphically divided by the highest point. Therefore, when looking at some specific year, it shows the search of the term in a given time period and the overall time period would have been assigned the peak of 100. The other time periods would be displayed as proportions of the volumes of searches for the peak period.

Normalisation of the data means that Google has divided sets of data by a common variable to cancel out the variable's effect on the data. This provides that the underlying characteristics of the data sets can be compared. This means that, for example, when looking at Google Trends data for two different locations, interest" (proportion of searches) rather than volume" is being compared [8,16].

Discussion

What is the interpretation of these data? The results of the research on the use of social networks and media in the online environment



using data-mining techniques (data mining) make it evident that entrepreneurs in tourism and tourism itself can benefit from Internet marketing by targeting individual offers to its customers, mainly with the use virtual travelling communities, sites with travellers' reviews and blogs. With the ever-increasing use of the Internet, the use of the methods mentioned in this paper will increase rapidly and the business of tourism not using these methods would economically suffer. The actual social networks, when searching for keywords using Google's search engine, played a minimal role. Use of data mining and correct interpretation of data obtained has helped to increase the awareness of the product or service in tourism, but not least significantly impact the growth in demand, direct targeting of supply and the effectiveness of marketing activities of entrepreneurs in tourism, non-profit organizations and government organizations. The importance of this paper might be seen beneficial by various entrepreneurs, local businesses, and other organizations in the tourism industry as it clearly defines which social media are worth being data-mined in order to get relevant outcomes in effectiveness, productivity and general awareness in the provision of their product and services to their target group tourists. We may also add that this paper serves as an insight into the current and innovative topic of usage the Internet and social media data mining for the tourist industry.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract no. APVV- 14 - 0797.

References

- Bose I, Mahapatra RK (2001) Business Data Mining –A Machine Learning Perspective. Information Management 39: 211-225.
- 2. Svoboda L (2013) Data Mining.
- Xiong H, Pandey G, Steinbach M, Kumar V (2006) Enhancing Data Analysis with Noise Removal. Knowledge and Data Engineering 18: 304-319.
- Balaz B, Strba L, Lukac M, Drevko S (2014) Geopark development in the Slovak Republic - alternative possibilities. Ecology and Environment Protection 2: 17-26.
- 5. Pearce LP (2005) Tourist Behavior Themes and Conceptual Schemes.
- 6. Blackshaw P, Nazzaro M (2006) Consumer-Generated Media (CGM).
- Schmallegger D, Carson D (2008) Blogs in tourism: Changing approaches to information exchange. Journal of Vacation Marketing 14: 99-110.
- Cehlar M, Domaracka L, Simko I, Puzder M (2016) Mineral resource extraction and its political risks. Production Management and Engineering Sciences 2: 39-43.
- Gretzel U, Yoo KH (2008) Use and impact of online travel reviews. In: Connor PO, Hopken W, Gretzel U (eds.) Information and communication technologies in tourism. Springer, New York.
- Vermeulen IE, Seegers D (2009) Tried and Tested: The Impact of Online Hotel Reviews on Consumer Consideration. Tourism Management 30: 123-127.
- Palomares J, Gutierrez J, Mínguez C (2015) Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. Applied Geography 63: 408-417.
- 12. Cibererova J, Korba P, Sabo J (2015) Innovative approach to developing e -learning.
- 13. Werthner, Ricci (2004) E-commerce and tourism 47: 101-105.
- 14. Albena, Bulgaria, Sofia (2014) STEF92 Technology Ltd.
- 15. Burns E (2007) U.S. search engine rankings and top 50 web rankings.
- Hopkins H (2008) Hit wise US Travel Trends: How Consumer Search Behavior is changing.
- Wober K (2006) Domain specific search engines. In: Fesenmaier DR, Wober K, Werthner H (eds.), Destination recommendation systems: Behavioral foundations and applications, CABI, Wallingford, UK.
- 18. Han J, Kamber M, Pei J (2011) Data Mining Concepts and Techniques.