

## Introduction: Special Issue on Business intelligence and big data in the travel and tourism domain

Wolfram Höpken<sup>1</sup> · Matthias Fuchs<sup>2</sup>

Received: 25 January 2016 / Revised: 17 February 2016 / Accepted: 19 February 2016 /  
Published online: 17 March 2016  
© Springer-Verlag Berlin Heidelberg 2016

Information and communication technologies (ICT) are widely adopted in the tourism domain and, thus, information on nearly all tourism transactions, customer needs and behaviour, and the complete tourism market structure are electronically available. The field of *business intelligence* deals with extracting such information from different data sources and source systems, analysing data by techniques spanning from interactive visualizations like OLAP to complex data mining methods, like machine learning or statistical approaches, and gaining new knowledge as input to decision support or intelligent and adaptive systems (like recommender systems or dynamic pricing).

Business intelligence constituted an important research field in tourism for more than a decade and even gained more attention with the advent of *big data*. Big data summarises trends, like integrating huge amounts of data from external data sources (e.g. web content), extracting information from any kind of data sources, especially unstructured data (e.g. customer reviews), and integrating data in real-time, when necessary. Business intelligence and big data are just about to unfold their full potential for the tourism domain. Due to the crucial role and importance of social media and online product reviews in tourism, the above trends become vital for tourism companies to ensure their competitiveness. More powerful IT systems and new algorithms and methods especially in the area of web content mining and text mining enable new application fields for business intelligence methods which have already gained high research attention.

---

✉ Wolfram Höpken  
wolfram.hoepken@hs-weingarten.de

<sup>1</sup> Business Informatics Group, University of Applied Sciences Ravensburg-Weingarten, Weingarten, Germany

<sup>2</sup> European Tourism Research Institute (ETOUR), Mid-Sweden University, Östersund, Sweden

The articles included in this special issue cover different facets of current research activities within the field of business intelligence and big data and give a comprehensive insight into current research questions and challenges in this fascinating research field.

The paper titled “*Analyzing User Reviews in Tourism with Topic Models*” by Marco Rosetti, Fabio Stella and Markus Zanker presents a novel approach for analysing customer reviews by making use of the unsupervised learning of topic modelling. Based on the topics mentioned by users within their reviews, the approach extracts user profiles, consisting of topic-wise user preferences, as well as item profiles, consisting of topic-wise customer satisfaction. The fit between concrete user profiles and item profiles then serves as input to a recommendation service. Additionally, the approach provides a multi-criteria characterization of user preferences and item ratings, enabling to offer a better explanation of a given recommendation, to compare different items or even to calculate ratings for given reviews. The contribution of the paper—the novel topic-criteria model and extended topic-sentiment-criteria model—have been evaluated based on a Yelp and TripAdvisor dataset, showing not only competitive or even superior accuracy, compared to existing approaches, but also demonstrating the practical relevance in the scenarios review analysis and interpretation, recommendation and rating calculation.

The paper titled “*Perceptual Mapping of Hotel Brands Using Online Reviews: A Text Analytics Approach*” by Matthew James Krawczyk and Phil Zheng Xiang uses a text analysis approach to create perceptual maps from the most frequent terms used in a data set collected from an online travel agency. The maps mirror the structure of hotel suppliers represented by hotel class and service attributes. These maps show the ability for text analytics to produce a deeper understanding of how hotel brands, comprising tangible and intangible assets, are formed and are able to be differentiated in the minds of consumers. Findings indicate that brands are associated by customers with terms indicative of the class of the hotel. The authors point to a new research direction theorizing about the nature of online reviews, particularly how online reviews may be used to understand the market structure of the hospitality and tourism industry.

The paper titled “*Automatic Analysis of Textual Hotel Reviews*” by Aitor Garcia-Pablos, Montse Cuadros and Maria Teresa Linaza presents a natural language processing (NLP) platform, providing text processing technologies like sentiment analysis, opinion mining and named entity recognition (NER), to be applied to textual content in the form of user reviews in the hospitality domain. The basic foundation of the presented approach is a homogeneous data representation format used along all different text processing steps. This so called knowledge annotation format (KAF) enables the provision of meta data on different layers for different text processing and analysis tasks and, thus, enables a flexible integration of different text processing modules. The presented approach has been evaluated on hotel reviews extracted from the review platforms Zoover and HolidayCheck, leading to satisfactory precisions of the different text processing tasks, and demonstrating its practicability in flexibly combining different tasks to an overall text processing pipeline.

The paper titled “*Towards Cross-domain Decision Making in Tourism: A Linked Data Based Approach*” by Marta Sabou, Irem Onder, Adrian M. P. Brasoveanu and Arno Scharl presents an approach for cross-domain data analysis based on an application of linked open data technologies to interlink statistical data from different domains in tourism. Based on ontologies for modelling statistical data (e.g. RDF Data Cube) as well as tourism-specific ontologies, data are semantically annotated and mapped to RDF resources in order to be available in the web in an open and flexible manner. A dashboard has been built upon such publicly available linked data from TourMIS, Eurostat and WorldBank, enabling cross-domain analyses and supporting tourism executives’ and researchers’ decision-making process on a broader data basis. Results of an evaluation by practitioners and domain experts indicate that requested information can be found more efficiently and analyses can be performed more precisely compared to traditional data analysis systems, based on separated data sources.

The paper titled, “*Discovering Social Influencers with Network Visualization: Evidence from the Tourism*” by Chiara Francalanci and Ajaz Hussain proposes a novel visual approach to the analysis and exploration of social networks. By exploiting the theory of k-shell decomposition, influencers (hub nodes) and influence (spread of multi-layer peripheral nodes), represented by the opinions of social media users on a given set of topics, are identified and visually highlighted. Results show that the approach produces aesthetically pleasant graph layouts, by highlighting multi-layered clusters of nodes surrounding hub nodes (i.e. main topics). Further empirical tests on the base of a large sample of tweets from the tourism domain show that the proposed hypotheses that tie content specificity, frequency of tweets and retweets are supported. Social media marketing managers can apply the proposed approach to analyse most competitive locations, events or initiatives in the market, thereby visually identifying key players in the network, like information spreaders and information sources.

The paper titled “*The Role and Impact of Comparison Websites on the Consumer Search Process in the US and German Airline Markets*” by Christopher Patrick Holland, Julia Andrea Jacobs and Stefan Klein illustrates a novel use of online panel data to explore detailed aspects of customer search behaviour, in particular the interaction effects between different types of web-sites, in this case comparison websites and airline websites. Set theory is employed to analyse audience duplication reports as a novel methodology. By using data from ComScore, the leading commercial provider of consumer analytics based on a worldwide panel of two million online users, the travellers’ search process is modelled using the concept of the consideration set based on primary search with the airline websites and the use of Online Travel Agents (OTAs) and meta-search engines, termed comparison websites. Study findings show that comparison websites are a catalyst for further direct research, rather than a substitute for direct search with individual airline websites.

We would like to thank all authors for their excellent contributions and all reviewers for the time they invested in reviewing the articles and giving valuable

input and stimulations. In addition, we would like to express our thanks to the Editor-in-Chief, Francesco Ricci, for his continuous support.

Guest Editors  
Wolfram Höpken  
Matthias Fuchs