Chapter 3
Big Data Analytics for Tourism Destinations

Wolfram Höpken
University of Applied Sciences Ravensburg-Weingarten, Germany

Matthias Fuchs
Mid-Sweden University, Sweden

Maria Lexhagen
Mid-Sweden University, Sweden

ABSTRACT

The objective of this chapter is to address the above deficiencies in tourism by presenting the concept of the tourism knowledge destination – a specific knowledge management architecture that supports value creation through enhanced supplier interaction and decision making. Information from heterogeneous data sources categorized into explicit feedback (e.g., tourist surveys, user ratings) and implicit information traces (navigation, transaction, and tracking data) is extracted by applying semantic mapping, wrappers, or text mining. Extracted data are stored in a central data warehouse enabling a destination-wide and all-stakeholder-encompassing data analysis approach. By using machine learning techniques interesting patterns are detected and knowledge is generated in the form of validated models (e.g., decision trees, neural networks, association rules, clustering models). These models, together with the underlying data (in the case of exploratory data analysis) are interactively visualized and made accessible to destination stakeholders.

INTRODUCTION

Although information and communication technologies (ICT) were an important issue for Travel & Tourism (T&T) since the 1960ies (i.e. computer reservations systems, global distribution systems; Werthner & Klein, 1999), the difference today is that ICT has become a strategic issue for every tourism business (Buhalis, 2006). The special benefit tourism gains from ICT can be put down to the characteristics of the tourism product, being a service bundle ideally portrayed by electronic media and being jointly de-

DOI: 10.4018/978-1-5225-7598-6.ch003

livered by (usually) small-sized enterprises. Indeed, T&T is a highly information intensive sector, and not surprisingly, T&T represents the largest branch within the e-Commerce sector. In 2012, 45% of the EU online sales volume has been generated by the T&T sector, whereat in 2008 this figure stood only at 25%. Moreover, in the US already 51.5% of the total travel revenue is generated online (E-Marketer, 2012). However, although tourism shows high penetration rates with respect to web-based marketing & distribution, shortcomings become evident with respect to e-business networks (supply-chains) and integrated process automation (e-procurement, enterprise resource planning, etc.). Finally, significant adoption gaps are ascertained for ICTs in tourism SMEs to support market research, product development and strategic decision making (e-Business Watch, 2006).

The attractiveness of tourism destinations particularly depends on how communication and information needs of tourism stakeholders can be satisfied through ICT-based infrastructures so that sustainable knowledge sources can emerge (Buhalis, 2006). Although huge amounts of customer-based data are widespread in tourism destinations (e.g. web-servers store tourists' website navigation, databases save transaction and survey data, respectively), these valuable knowledge sources typically remain unused (Pyo, 2005). However, managerial effectiveness and organisational learning could be significantly enhanced by applying methods of *business intelligence* (BI) and *big data analytics* (Wong et al., 2006; Shaw & Williams 2009), offering reliable, up-to-date and strategically relevant information, such as tourists' travel motives and service expectations, information needs, channel use and related conversion rates, occupancy trends, quality of service experience and added value per guest segment (Min et al., 2002; Pyo et al., 2002). This makes clear why ICT and methods of BI are playing a crucial role in effectuating a *knowledge destination* by enhancing large-scale intra and inter-firm knowledge exchange. Indeed, the major challenge of knowledge management for tourism destinations is to make individual knowledge about customers, products, processes, competitors or business partners available and meaningful to others.

The objective of this chapter is to address the above deficiencies in tourism by presenting the concept of the tourism knowledge destination – a specific knowledge management architecture that supports value creation through enhanced supplier interaction and decision making. Information from heterogeneous data sources categorized into explicit feedback (e.g. tourist surveys, user ratings) and implicit information traces (navigation, transaction and tracking data) is extracted by applying semantic mapping, wrappers or text mining (Lau et al., 2005). Extracted data are stored in a central data warehouse enabling a destination-wide and all-stakeholder-encompassing data analysis approach. By using machine learning techniques interesting patterns are detected and knowledge is generated in the form of validated models (e.g. decision trees, neural networks, association rules, clustering models). These models, together with the underlying data (in the case of exploratory data analysis) are interactively visualized and made accessible to destination stakeholders. The technical architecture and implementation issues are discussed based on a prototypical implementation for the leading Swedish tourism destination, Åre (Höpken et al., 2015).

BACKGROUND

Since the widespread adoption of computerized reservation and booking systems in the 1980ies, comprehensive databases are available for all types of tourism transactions, i.e. the complete booking and consumption behavior (e.g. Passenger Name Record (PNR) databases of global distribution systems (GDS) or the airline on-time performance database of the Bureau of Transportation Statistics; BTS, 2012). Immediately, especially airline companies started to analyze such data as input to process and