Semantic Social Media Analysis of Chinese Tourists in Switzerland

Zhan Liu[1], Jialu Shan[2], Nicole Galassey Balet[1], Gang Fang[3]

[1]Institute of Information Systems
University of Applied Sciences and Arts Western Switzerland (HES-SO Valais-Wallis)
Sierre, Switzerland
Email: zhan.liu@hevs.ch

[2]Global Center for Digital Business Transformation
International Institute for Management Development (IMD)
Lausanne, Switzerland
Email: jialu.shan@imd.org

[3]School of Management
Hangzhou Dianzi University
Hangzhou, China
Email: fanggang@hdu.edu.cn

Abstract

In recent years, Sina Weibo, a Twitter-like social network service in China, has attracted attention from scholars in the domain of information systems, as the spread and influence of users' opinions are increasingly important, particularly in the tourism industry. This study examined the behaviors of Chinese tourists in Switzerland by adopting a semantic-based linked data methodology. A total of 103,778 Weibo messages shared with Swiss locations were collected between January 2013 and April 2015. We addressed questions about Chinese travelers' profiles, trends in keywords, and differences between first time and repeat visitors. Moreover, we implemented a semantic search engine by employing linked data technologies to provide useful information about Chinese tourists in Switzerland, both for the tourism industry and individual tourists.

Keywords

Semantic Web, Linked data, Social media, Tourism, Big data, Chinese tourists

1. Introduction

In the past decade, with the popularity of online social networking services (SNS), such as TripAdvisor, Facebook, and Weibo, more and more data have become available to the tourism industry, as people tend to leave different digital traces on their social networks when travelling. However, tourism-related big datasets often originate from various sources and are difficult to compare. To overcome these problems, big data mining including that of semantic web-based linked data, has been developed to support the intelligent integration of tourism data. As a method to connect related data on the web, linked data are important in improving the quality of natural language processing (NLP), such as users' opinions found on social media platforms. By evaluating and analyzing large amounts of data on travelers' opinions of hotels, restaurants, destinations, and so forth, the results can deliver more consistent business insights, and can be a source of innovation for the tourism industry and related scientific research (Fuchs et al. 2014; Xiang et al. 2015). Nevertheless, to date, only a small number of related studies on this subject has been published in the tourism literature (Francalanci and Hussain 2016; Garcia-Crespo et al. 2011).

China, with its double-digit growth in expenditures every year since 2004, leaped to first place in international tourism expenditures in 2012. In 2015, it consolidated its position as the number one tourism source market in the world, spending US $165 billion on international tourism, a 28% growth rate from 2014. Switzerland, as one of the most desirable European destinations for Chinese tourists, has been a strong and consistent growth market during past years. Taking advantage of increasingly fewer restrictions on travel visas, as well as an appreciating currency, Chinese consumers will continue to make significant contributions to the Swiss economy (Switzerland Tourism, 2013). A better understanding of Chinese consumers is therefore essential to successful business practices. Chinese tourists today increasingly want to travel to foreign countries by themselves or in small groups of friends. In deciding their destinations, they often turn to online user-generated content (UGC) available through social media to obtain information prior to their vacations, believing that other tourists' experiences are trustworthy and useful (Yoo et al. 2009). Sina Weibo (microblog)—the largest and leading social media platform in China—now has over 600 million users. Weibo's large market penetration suggests that tour operators and the industry overall need to understand ways in which to build effective and sustainable communications on this Chinese social media platform. However, few studies have examined actual outbound Chinese travelers' attitudes and experiences empirically (e.g., on social media platforms) through the methodology of semantic-based big data analysis. Research on data linked with behavior analysis to support the tourism decision-making process is also notably lacking.

To this end, the goal of this research was to identify and understand Chinese tourists' behaviors in Switzerland by adopting a linked data approach on Chinese social media platforms. We conducted a large-scale big data analysis with semantic-based linked data and explored Chinese tourists' attitudes about, and behavior in Switzerland, as expressed on Sina Weibo. In particular, we built an automatic multilingual semantic annotation system associated with the integration of BabelNet ontology (Navigli and Ponzetto 2010). All of the concepts extracted from Weibo messages, including related links and semantic interpretations, were stored in a graph-based RDF triple store to answer SPARQL queries through visual interfaces. We combined and applied both tasks of Entity Linking (EL) and Word Sense Disambiguation (WSD) in our methodology to achieve the research objectives, and finally, more accurate results were obtained by using WSD and automatic semantic annotation approaches together.

Thus, we addressed the following three research questions in this study:

- *Who are these tourists and where do they travel in Switzerland?* Because real name registration exists in Weibo, it is possible to identify these tourists' characteristics, such as gender, age, etc. As we looked at Sina Weibo messages that shared location information with that in Switzerland, the analysis was more likely to provide unbiased results.
- *What are the new trends among Chinese tourists in Switzerland?* For example, how are their travel habits in Switzerland changing and what are their effects on tourism destinations?
- *How does the semantic-based methodology help us achieve a better understanding of traveler's experiences with Swiss tourism?* With the help of linked big data mining, it was possible to determine

whether or not they were first time or multiple time visitors. We compared these two different types of visitors and explored in what ways first time visitors are similar to or different from multiple time visitors in order to determine how their behavior develops over time. Moreover, we developed a semantic-based social search engine, which can be used as a suitable method to help destination managers find, recognize, link, and manage resources related to Chinese tourists' travel experiences in Switzerland.

2. Background and Related Work

In this section, we explore related work on the characteristics of social media data and then present the effects of social media on user behavior. Finally, we highlight some of the methods we used to analyze social media data with the Semantic Web.

2.1. Big Data and Social Media

Big data are defined as those that include several attributes, not simply high-volume data. According to Forrester (2012), big data include four main attributes: (1) volume—approaches or exceeds physical limits of vertical scalability; (2) velocity—small decision window compared to data, and data requirements change rate; (3) variety—many different formats that make integration expensive, and (4) veracity—messiness or trustworthiness of the data. Marr (2015) suggested that value is an important point, and should be considered the fifth V when describing the phenomenon of big data. He argued that, in essence, when talking about big data, it refers simply to vast amounts of data that are potential treasure troves of information. However, the process of analyzing these data is equally important in converting data into value. A recent Forrester's Business Decision-Makers survey conducted in 2012 with 3,616 business decision-makers from over 100 firms worldwide deepened this understanding. These executives indicated that the three most important aspects of big data are their ability to improve the use of data and analytics to enhance business decisions and outcomes, improve IT project delivery performance, and develop new skills to support emerging technologies and business innovation better. The McKinsey Global Institute described big data as the next frontier in innovation, competition, and productivity (Manyika et al. 2011). However, as much as 90% of the data is unstructured, meaning that it is generated spontaneously, and is not captured and classified easily, which is reflected particularly in the domain of social media networks.

Because of the popularity of various online SNS, such as Facebook, Twitter, and Weibo, they have become a major source of big data. Indeed, SNS have become the most effective and fast channels of information diffusion and dissemination (Shen et al. 2015). Online social media big data have been hailed as key to crucial insights into human behavior and have been analyzed extensively by scholars, corporations, politicians, journalists, and governments (Doyd and Crawford 2012; Lazer et al. 2009). SNS contain rich information about people's daily lives, including their activities, interests, locations, shares, and likes. As a result, SNS not only have changed the way we communicate, but also have affected how we view the people around us and ourselves.

Use of the Internet as a means of interpersonal communication has grown dramatically in China in recent years. Statistics show that the Chinese spend an average of 90 minutes per day on social media sites. Fifty-eight percent of all Chinese Internet users use social networks, and among these, 88% are active on at least one social media network (Makeawebsitehub 2015). The most popular social media platform in China is Sina Weibo. Launched on 28 August, 2009, Sina Weibo has grown rapidly, and by the end of 2015, it had reached 600 million registered users, with 222 million active monthly users (China Internet Watch 2015). The current volume of Weibo messages is over 2.8 billion per month. Sina Weibo users tend to be young—the vast majority is under the age of 32 years—as well as better educated and wealthier than is the general population. By providing a simple way of communicating, Sina Weibo allows people to express themselves and be heard publicly.

2.2. SNS Users and Use Behavior

As social media have flourished, the analysis of user behavior in SNS has become an increasingly important topic. A substantial amount of research has been conducted in recent years on user demographics and their specific SNS behaviors. For example, youth are more likely to spend more time on SNS and view and share personal information, such as photos, than are older adults (e.g., Zhang and Pentina 2012). Research also has suggested that gender plays an important role in the online behavior of SNS users. Muscanell and Guadagno

(2012) found that men tend to use SNS to form new relationships, while women use them to maintain relationships. This is consistent with women's general tendency to use the Internet to communicate interpersonally, while men are more likely to spend their time online engaging in more task-related activities, such as reading news, etc. Such a phenomenon also exists in online social networking use. While statistics have indicated that there is no obvious gender difference in the overall amount of Sina Weibo use (Weibo's User Development Report in 2014), there are gender differences in the motivations for its use and the use of time spent on it. Hwang and Choi (2016) found that female students are more likely to use Sina Weibo in a much broader manner, accessing it to satisfy all manner of needs, including information gathering, accessibility to celebrities, social connection, self-presentation, and entertainment.

Regardless of the demographic differences, consumers in the travel context often disseminate Word of mouth (WOM) because travelling is not associated with a product, but rather an experience (e.g., Litvin et al. 2008; Burgess et al. 2009; Cox et al. 2009; Jansen et al. 2009). This is particularly true when planning travel to a destination they have not visited previously (Litvin et al. 2008). In the social media era, consumers use SNS intensively for tourism and travel-related products. According to Yoo and Gretzel (2010), approximately half of their respondents read other travelers' posts when they planned their most recent overnight pleasure trips, and the majority (81.6%) trust the information posted by other travelers. Simms (2012) found that a higher percentage of tourists turn to UGC when visiting a destination for the first time, as well as when visiting an international destination. Jacobsen and Munar (2012) also showed that when tourists to a well-known destination decide where to spend their summer holidays, they tend to use social media. Thus, UGC generated by SNS has become one of the most influential sources of information for travel decision-making.

Recently, as Sina Weibo has become an increasingly popular platform through which Chinese Internet users communicate, the academic world has witnessed rapid growth in research on UGC-based analysis of data on Sina Weibo. Chua et al. (2012) proposed a framework to index, organize, and search for live and dynamic UGCs on social media to generate higher-level analytics that can be used to understand topics of interest to society, sentiments, evolving live events and social communities, as well as fashion habits and trends. Some studies have focused on whether UGC can be used to identify or predict user behavior. For example, Bai et al. (2013) proposed using an incremental regression model to analyze UGC in order to predict Weibo users' personalities and behavior. The results demonstrated the feasibility of their approach, verified the reliability of their dataset, and concluded that UGC is important in reflecting users' online behaviors. Moreover, Li et al. (2015) studied user behaviors related to topics popular on Weibo during August 2009 and January 2012, and found that UGC and reposting behaviors were associated with demographic factors, including gender, verification status, and regional location.

Despite the growing amount of research that has investigated UGC and SNS in the tourism field, there exists little information to date about Weibo user behavior with respect to Chinese tourists' expectations and experiences in Western destinations. Understanding the way in which linked data technologies can help generate insights about the use of SNS for travel-related purposes is also limited. Hu et al. (2013) conducted an analysis of pictorial tourism content on Sina Weibo, regarding "Switzerland Travel" as a tourist destination in an attempt to understand the most important aspects of Chinese tourists in Switzerland. However, we believe that there were some biases in their study: first, the data used in the study were collected and analyzed through human interpretation, and as a result, only Weibo posts that contained original pictures were included in the final analysis. However, according to the Weibo User Development Report in 2014, more than 30% of Weibo messages do not contain pictures. These messages may provide equally valuable insights about Chinese travelers' perceptions of Switzerland. Moreover, in June 2014, a query of Weibo's search engine using the keywords "Switzerland Travel" returned nearly 2 million results. Analyzing and interpreting such a huge amount of information via any human endeavor is futile. Second, and more importantly, the keywords "Switzerland and Travel" failed to be good representatives, because Weibo users may post text messages without mentioning the words "Switzerland" or "travel" explicitly, although they are referring to the subject of travel in Switzerland. For example, a Weibo message such as, "I am enjoying fondue in Valais" or, "I am in front of the Olympic museum in Lausanne" are related clearly to travel experiences in Switzerland. In addition, at the time the data were collected in June 2012, there were fewer than 80 million active monthly users, with 3.5 million active daily users.

Both numbers almost doubled in early 2014. Therefore, it is necessary to conduct new studies to understand Chinese tourist behavior and trends with respect to Switzerland.

2.3. Semantic Annotation of Social Media

Mining UGC with semantic technologies has received increased interest from researchers and practitioners, but is still in its infancy. The need to understand the content of social media users' messages more fully has led to improved algorithms to process natural language. The process of combining semantic models and natural language is referred to as semantic annotation, which allows dynamic and bidirectional exploration of the interrelationships between ontologies and unstructured multilingual documents (Kiryakov et al. 2004). Consequently, semantic annotation involves annotating all mentions of concepts in texts from the ontologies, including classes, instances, properties, and relationships, through metadata that refer to their URIs in the ontology (Bontcheva and Cunningham, 2011). Semantic annotation can be performed manually or automatically; the former is usually accompanied by post-editing and correction by human annotators. For example, Hepp (2010) proposed a manual semantic annotation syntax for Twitter messages to map with RDF statements. The syntax provided relationships between tags/properties and multiple RDF statements in the same tweet message. Passant and Laublet (2008) introduced the Meaning of a Tag ontology to define the semantic meaning of tags by linking open data to perform annotations of social media manually. In addition, Djuana et al. (2011) and Pöschko (2011) proposed a text classification by using WordNet ontology from user tagging information to represent the semantic meaning and hierarchical relationship among tags. Finally, Vicient and Moreno (2015) presented a topic discovery methodology that uses WordNet terms to map hashtags manually according to concepts of knowledge-based semantic similarity.

While such manual semantic annotation efforts are valuable, automatic semantic annotation methods are required to make sense of the millions of messages on popular SNS. Given the large number of messages posted on Weibo daily, we focused on automatic approaches in this study. Ontology-based information extraction is adapted specifically for the automatic semantic annotation of UGC (Li and Bontcheva, 2007). It refers to model tagging, sharing, geographical location, and multilingual text messages from social media that enable semantic-based searches, browsing, filtering, and visual analytics. In these approaches, the ontologies can serve not only as system input, but also as the target output (Bontcheva and Rout, 2014). Linked open data resources, such as BabelNet, DBpedia, and GeoNames have become key sources of ontology and knowledge bases for semantic annotation and disambiguation. By linking DBpedia and GeoNames ontologies, Kleinen et al. (2014) provided a mobile application for retrieval, search, and exploration of resources, such as places and events on social media. Kobilarov et al. (2009) described the way in which the public service broadcaster works to integrate data and link documents using Semantic Web technology, in particular linked data and DBpedia. Moreover, BabelNet is constructed automatically through Semantic Web technologies that establish a map between a multilingual encyclopedic knowledge repository (e.g., Wikipedia) and a computational lexicon of English (WordNet), to construct a large multilingual lexical knowledge resource automatically (Navigli and Ponzetto 2010).

To summarize, automatic semantic annotation is a relatively new area in social network analysis and research in the field is still evolving. However, most existing studies focus on mining social data in the English language, and very little research has been conducted using semantic technology-based data mining in Chinese social media. Therefore, we designed this study to fill that research gap.

3. Methodology

3.1. Sample Description

This study analyzed Weibo messages from Chinese tourists with respect to their expectations and real experiences in Switzerland to gain insights into the tourists' profiles and their Weibo use patterns. The objectives, therefore, were not simply to determine who the bloggers are and when or where they blog. More importantly, we wished to determine what they said and identified new trends in their travels in Switzerland.

To do so, our goal was to collect all Swiss or Swiss travel-related Weibo messages. The most efficient method of data acquisition is to use an application programming interface (API), because the data are returned in a data-interchange format, JSON, which is easy for humans to read and write. However, there were more than 248 million Swiss-related messages on Weibo by the end of 2015. On the one hand, the limit in the number of single

account requests reduced the possibility of data collection through Weibo API significantly: it would take several years to acquire such a huge number of Weibo messages. The streaming API provided by Sina Weibo, on the other hand, allows access only to certain information contained in Weibo messages, such as text, timeline, comments, replies, and reposts, but it does not allow the Weibo messages to be filtered geographically, nor can any user profile data be collected.

Due to these restrictions on data acquisition, the final data were collected and provided by our strategy partner— Weibo Data Center. Each piece of data included the multilingual Weibo message (i.e., Chinese, English, French, and German), post date and time, posting address with latitude and longitude, Weibo user's name, the number of reposts, comments, and likes, as well as the posters' profiles, including name, gender, location, birthday, name of school, description, account creation date, and the number of followers, friends, and Weibo messages. To achieve our research objectives, we focused on messages in Sina Weibo's database that included Swiss geographical information. We excluded data from users whose registered home address was in Switzerland, as they could not be considered "tourists." As we also intended to explore the trends in Chinese tourists' behaviors, the data were collected over a three-year period from the beginning of January 2013 to the end of April 2015.

3.2. Semantic-based Natural Language Data Mining

Traditional natural language data mining and search methods are no longer satisfactory to address the more complex ways in which users intend to seek information in social media, which involves sense making, investigation, and social search (Pirolli 2009). Semantic technologies have the potential to help us handle the massive amounts of social media-based information more effectively. The greatest challenges in analyzing the characteristics of social media are focusing on noisy content, social context, user-generated resources, and multilingual texts (Bontcheva and Rout 2014), which provide opportunities for the development of semantic-based linked data approaches.

Compared to most existing studies (Xiang and Gretzel 2010; Miguéns et al. 2008; Cheng and Edwards 2015), we analyzed multiple languages based text messages in this research, including Chinese, English, French, and German. We adopted the Semantic Web-based Linked Open Data (LOD) technology to understand what Chinese tourists talk about during their travels in Switzerland. LOD is believed to be a promising approach to process natural languages interoperably based on lexical ontologies, such as BabelNet (Navigli and Ponzetto 2010), Wikipedia (Gabrilovich and Markovitch 2007), and WordNet (Miller 1995). The goal of this approach is to link and identify relevant concepts and entities described in users' Weibo messages by EL and WSD. EL is able not only to identify named entities, but also can manage cases in which an entity is mentioned only in part (e.g., in a case in which only a person's surname is given, EL can identify the person). WSD is used to address homonyms and uses the context to choose the correct sense of a word with different meanings. These two tasks are similar; the primary difference between them is that in EL, the textual reference can be linked to a named entity that may or may not contain the exact reference, while in WSD, there is a perfect match between the word form and suitable word sense.

Babelfy (Moro et al. 2014a) is a novel, unified, graph-based approach to WSD and EL that is able to reweight the edges of a large semantic network. According to Moro et al. (2014b), Babelfy achieved better performance in BabelNet than did WordNet and Wikipedia. Recently, it has been used widely to link data and process natural language. Therefore, we implemented Babelfy 1.0 and BabelNet 3.5 in this study. As shown in Figure 1, all Weibo messages from our sample were sent first to the semantic network-based Babelfy API to perform disambiguation and EL. In this step, we extracted and removed all noisy content from the original resources. In fact, social media content often contains unusual spelling, irregular capitalization, and idiosyncratic abbreviations, and the EL approach helped us normalize spelling and capitalization. Moreover, it allowed automatic language detection to separate the Weibo text messages in language clusters for processing with different algorithms. Then, one list of semantic annotations that consisted of the concept identified by BabelNet, which contains the resources from DBpedia and GeoNames, and their corresponding positions in each Weibo message, were generated from Babelfy. We stored all annotations for 3,230,461 statements in RDF format and loaded them in the GraphDB referred to as SWICICO triple store for further retrieval and manipulation with SPARQL (Pérez et al. 2009). Subsequently, DBpedia identified a total of 16,677 concepts (Auer et al. 2007), including multilingual labels and definitions, thumbnail images, and information on location concepts from the

geographical ontology GeoNames (Wick and Vatant 2012). These linked data resources provided millions of instances of rich, cross-referenced, domain-independent entities named hierarchically, as well as their links to DBpedia, GeoNames, and other external data. In addition, an inter-linked and complementary set of resources with synonymous lexicalizations was created to identify users' social contexts (e.g., how frequently they interact, who follows the user, where they prefer to post, which users are similar) and user-related resources (e.g., gender, location, age, date, interests, and opinions).
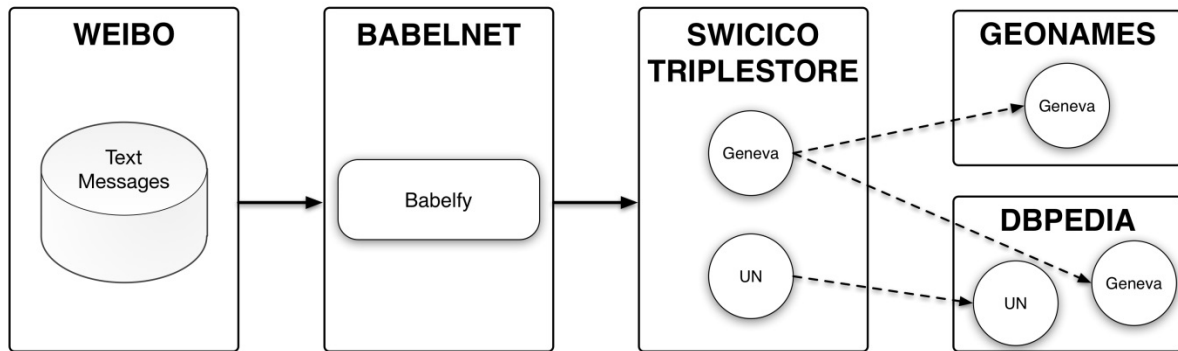
Fig. 1. Weibo data mining process with linked data technologies

4. Analysis and Discussion

Unlike most existing studies, in which data collection has focused on relevant keyword searches, such as "Switzerland travel," we relied on users who shared their location information. As mentioned previously, one of the main problems in using keyword searches is that such data are not very representative. On the one hand, Weibo users often post text messages to share their travel experiences without mentioning the word "travel" explicitly; on the other hand, there are huge numbers of text messages that mention the word "Switzerland" and "travel" but actually are talking about the poster's intention to visit Switzerland. As a result, the dataset contains too much noise that threatens the validity of any conclusions. Therefore, for the purposes of this study, we investigated all of the related Weibo messages with a Switzerland location tag.

4.1. Who Are They?

The 103,778 Weibo messages collected were written by 40,708 travelers while they were in Switzerland, as some tourists wrote more than one Weibo message per trip, and some visited Switzerland more than once during the data collection period. Approximately 70% of those 40,708 users were female (Figure 2). During the past three years, the percentage of female users has increased almost 10%. Therefore, we can assume that Weibo blogging is biased towards females.
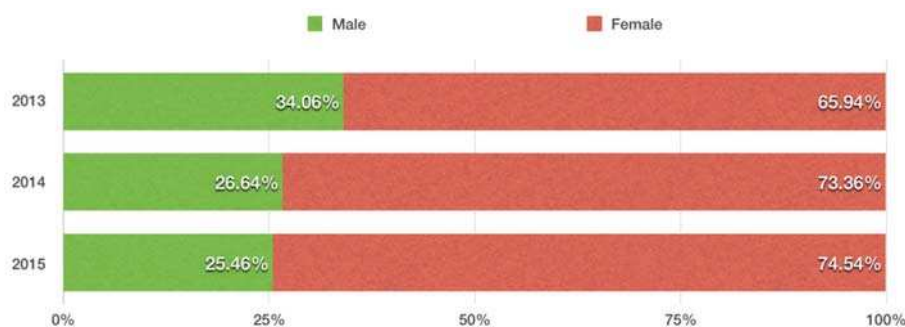

Fig. 2. Gender trends of Weibo users in 2013–2015* (when travelling in Switzerland)
Notes: *until April 2015

Consistent with previous findings, the age profile of Weibo users travelling in Switzerland tended to be young: the average age of our sample was 25.8 years (SD=7.53). Over 87% of the total Weibo users are between 16 and 35 years old. However, there is a limitation in profiling Weibo users demographically, as they do not always

provide their age, and approximately 30% of those in our sample did not do so. However, the remaining 70% suggested that Weibo users have tended to become younger in recent years. The largest age group in 2013 was 26–35 years old, while in 2014, the largest group was 16–25 years old (Figure 3).
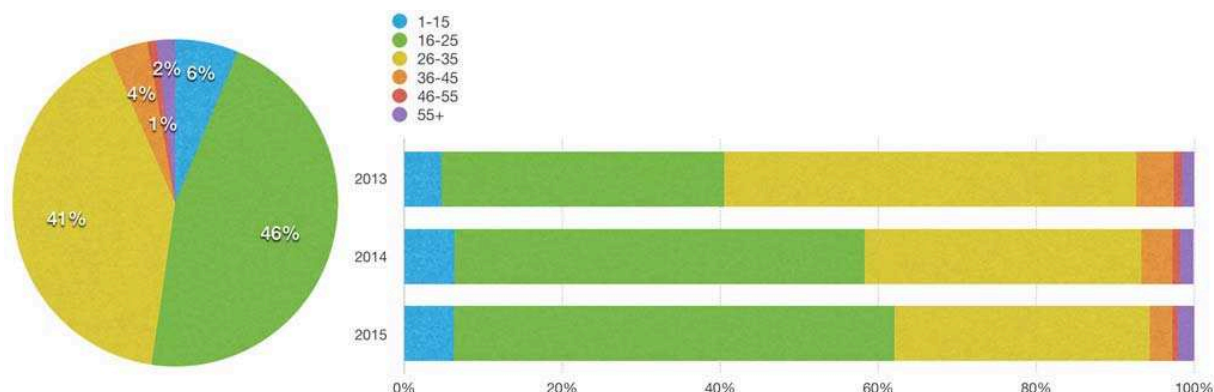

Fig. 3. Age trends of Weibo users in 2013–2015* (when travelling in Switzerland)
Notes: *until April 2015

As China is a huge country, and different regions usually have quite different backgrounds and cultures, we were also interested in determining where these tourists came from. Therefore, we conducted a geographical analysis at the level of province/direct-controlled municipality (e.g., Beijing, Shanghai). It is interesting to note that quite a few users registered as originating in Hong Kong, Taiwan, and overseas. We excluded these users for two reasons: first, many people might register their location as overseas simply for fun, and second, and more importantly, our focus was on the regional distribution of travelers from mainland China.

Figure 4 shows the top 15 regions of origin for our sample of Chinese travelers in Switzerland. It is clear that Beijing, Shanghai, and Guangdong were the top three regions of origin. Other high-ranking provinces included Jiangsu, Zhejiang, Sichuan, Chongqing, and Tianjin, all of which are economically developed and populous. The trend analysis showed that although Beijing, Shanghai, and Guangdong have remained the top three regions for the past three years, their percentage is declining, which indicates that more and more Chinese travelers from other regions are travelling to Switzerland, reflecting the ongoing diversity in China's outbound market.
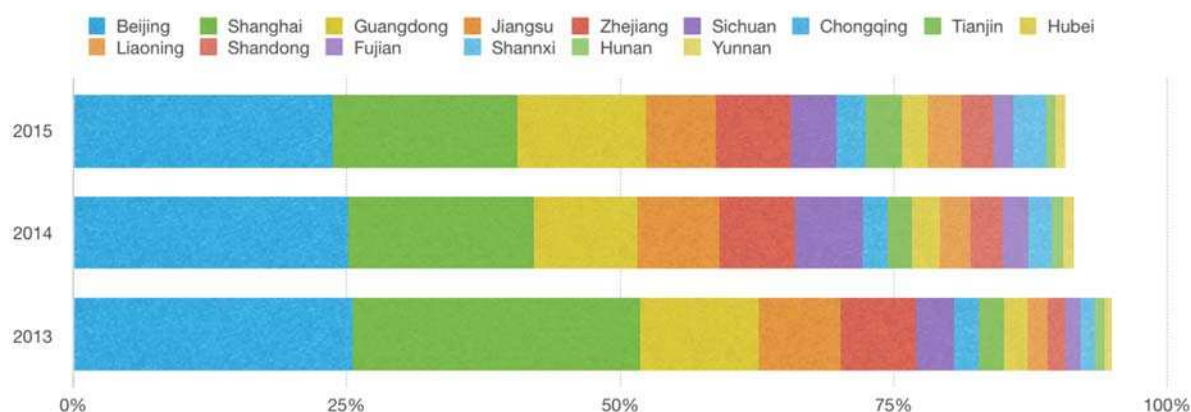

Fig. 4. Regional trends of Weibo users in 2013–2015* (when travelling in Switzerland)
Notes: *until April 2015

4.2. Trending Keywords

Sina Weibo provides a list of 50 keywords that appear most frequently in users' messages while travelling in Switzerland over a certain period and updates it constantly. Unlike most previous studies, which have examined the keywords offered by Sina Weibo (e.g., Yu et al. 2011; Guan et al. 2014), we investigated the top 30 noun keywords that appeared in our sample. Because there is a high rate of reposting about popular events, which

introduces some noise when calculating frequencies, we focused on original posts only, and ranked the top 30 nouns according to their frequency of appearance in our sample for the years 2013–2015, respectively (Table 1).

We found that some keywords appeared only in 2015, including *WeChat*, *Montreux*, *watch*, and *school*. As a new emerging message application, WeChat has become China's leading social interaction application and is becoming increasingly popular worldwide. As compared with Sina Weibo, WeChat is more personal and people prefer to leave their WeChat account ID in their text messages to discuss more details about their travelling related places, activities, and products with those who have common interests. We observed that Weibo's ranking, although still higher than that of WeChat, has decreased gradually, from 11th in 2013 to 17th in 2015. *Montreux* is a resort town on the eastern end of Lake Geneva in the Canton of Vaud that is famous for its beautiful lake view and the Jazz Festival held there each July. The castle of Chillon is another popular sightseeing attraction nearby. For these reasons, travel agencies generally exclude Montreux when they arrange large group tours. However, our findings suggested that this situation is changing. The other two new keywords, *watch* and *school*, indicated that Chinese travelers are becoming more interested in Swiss watches and education systems.

We also found that some keywords received more attention than did others. For example, Lucerne's rank increased from 14 to 9, while, in contrast, Nature received less attention. Another interesting observation is that the rank of some keywords tended to decrease in 2014, but increased again in 2015. One good example is *snow*, which ranked 6th in 2013, dropped to 10th in 2014, and then returned to 8th in 2015. One likely reason is that there was much more snow in 2013 than in the other two years.

TRENDS	2015	2014	2013	NOUNS
→	1	1	1	Switzerland
→	2	2	3	Snow Mountain
↗	3	6	8	Zurich
↗	4	5	7	Geneva
→	5	4	5	Town
→	6	7	4	Hotel
↗	7	9	13	Feeling
→	8	10	6	Snow
↗	9	11	14	Lucerne
↗	10	12	15	Alps
↗	11	18	17	Weather
→	12	14	10	Jungfrau
↘	13	8	9	Train
↗	14	16	16	Interlaken
↗	15	17	19	Lake
↘	16	15	12	Landscape/Nature/Scape
↘	17	13	11	Weibo
↗	18	20	19	Mountain
↗	19	24	27	Chocolate
↗	20	22	-	Basel
✿	21	-	-	WeChat
↗	22	25	20	Bern
↗	23	30	21	Breakfast
↗	24	29	24	Restaurant
→	25	26	23	Lausanne
✿	26	-	-	Watch
↗	27	28	22	Richness
✿	28	-	-	Montreux
↘	29	21	25	Swan
✿	30	-	-	School

Table 1. Top 30 nouns that appeared in Swiss travel-related Weibo messages (based on location information) in 2013-2015* Notes: *until April 2015

4.3. First Visit vs. Multiple Visits

To separate Weibo users who were visiting Switzerland for the first time from those who have visited more than once, we created a semantic-based algorithm that focused on users' social context to calculate whether their

posting interval spanned more than one month, and clustered similar users into different groups. Ultimately, we classified approximately 200 users as having visited Switzerland on multiple occasions.

To explore the Swiss cities Chinese tourists visited most frequently, we put their message locations on a map of Switzerland and separated them into two groups, first time and multiple time visitors. As shown in Figure 5, each light dot represents one Weibo message sent from that location by a first time visitor, while the dark dots represent messages from multiple time visitors. This revealed that first time visitors preferred to visit the most famous cities and places, such as Zurich, Lucerne, Interlaken, Bern, Lausanne, and Geneva. Multiple time visitors also like to go to these popular cities, but they are interested in visiting other attractions as well, as shown by the dark dots in Figure 5. One possible reason for this is that Chinese travelers who visited Switzerland more than once were more likely to go on an individual tour, rather than a group tour with a travel itinerary that covers only the major tourist attractions. Our findings could help tourism-related companies (e.g., hotels, restaurants, souvenir shops) understand and meet their customers' needs better by allowing them to provide high quality services to tourists. They also allow tourism destination managers to identify key factors that influence tourists' travel decisions and obtain new insights that drive significant business value.

Fig. 5. Regional distribution of first time (light dots) vs. multiple time (dark dots) visitors in Switzerland

By analyzing the content of travelers' Weibo messages while in Switzerland, we identified several key themes, as shown in the dark circle on the right in Figure 6. Similarly, we performed the same content analysis for first time visitors and collected the most frequent themes mentioned in their Weibo messages, shown in the light circle on the left in Figure 6. Interestingly, we found several themes that intersected. For example, Weibo messages from both groups mentioned cities such as Geneva, Lucerne, Zurich, and Bern frequently. All Chinese travelers also were interested in talking about chocolate, landscapes, snow-covered mountains, and castles. However, popular themes among multiple time visitors also included city names such as Interlaken, Lausanne, Montreux, and St. Moritz, while first time visitors' messages mentioned no other new city names. This finding is consistent with the geographical analysis that showed that multiple time visitors tended to post their Weibo messages in different regions.
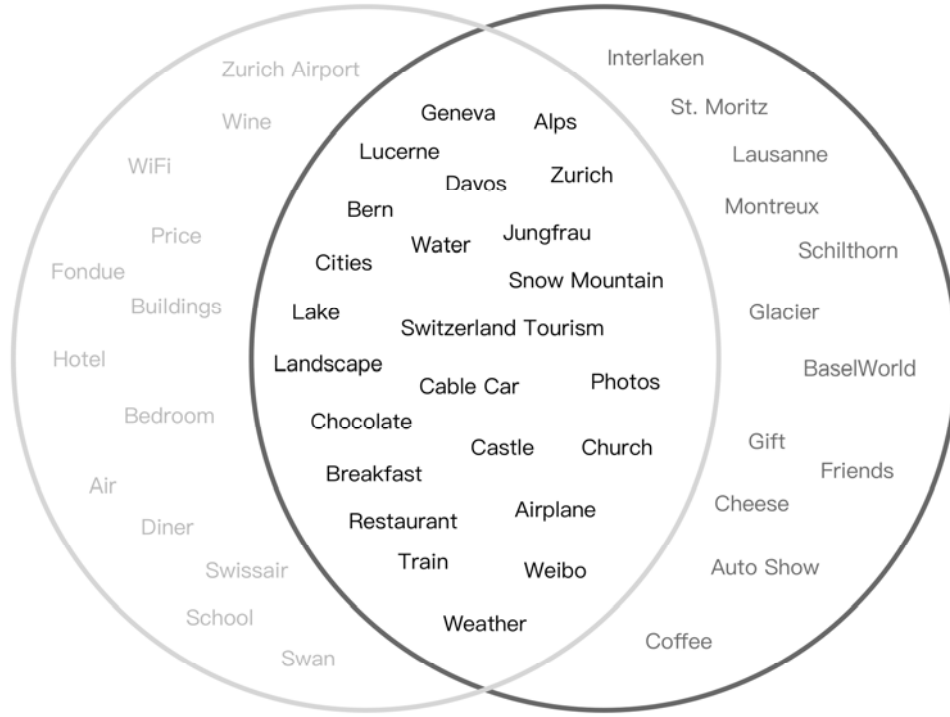
Fig. 6. Comparison of key themes for first time (light circle) vs. multiple time (dark circle) visitors in Switzerland

4.4. Semantic Search Engine

Given that our triple store contains a large number of semantic annotations of concepts in different domains, we implemented a semantic search engine by employing a linked data technologies-based approach. The goal was to provide not only keyword searches for matching text, but also to discover the linked concepts that Chinese tourists use in their Weibo messages. Moreover, the results provided by our search engine could serve both the tourism industry and individual tourists. On the one hand, tourism industry managers could use this linked information to explore potential services further, and determine the best combination of products based on tourists' experiences over time. On the other hand, the search function provides useful information to individual tourists making decisions during their travels. For example, they can find the most famous places to ski, and the restaurants and shops recommended most highly.

As described in the methods, we identified 16,677 concepts and linked them to knowledge bases including BabelNet, DBpedia, and GeoNames. Further, Weibo contains multilingual text messages. Most are written in Chinese (90%), although others include English, French, German, and Italian. Semantic technology methods have focused primarily on English to date, while their adaptation to multilingualism remains an open issue. To achieve this goal, we applied Babelfy API to build an automatic multilingual semantic annotation system, as it is the most accurate approach to processing natural language in Chinese. When a keyword is entered in the search box, a graph is created with the chosen concept in the center and related concepts surrounding it. As shown in Figure 7, when "skiing" is the keyword, ten related concepts, including some city names, appear around it. These are the top 10 other concepts that appear in Weibo messages that contain the concept "skiing" in multiple languages. As the concept was linked and defined through URI from the ontologies, this provides the ability to obtain the multilingual names and corresponding definitions of the concept. Therefore, if a traveler is interested in skiing in Switzerland, s/he can find valuable information by clicking on surrounding concepts that will provide him/her with detailed information about the concept offered by BabelNet, DBpedia, and GeoNames. In

11

addition, users can see the Weibo messages that contain both the keyword "skiing" and one related concept, i.e., co-occurrences, so that they can see what other travelers posted on Weibo.

The semantic search also provides an option to switch focus. For example, if a traveler decided to ski in Zermatt, s/he could switch the focus to list Zermatt as the keyword; another ten related concepts would appear around the new center. In this way, such information helps travelers find other tourist attractions near Zermatt when they plan a trip.
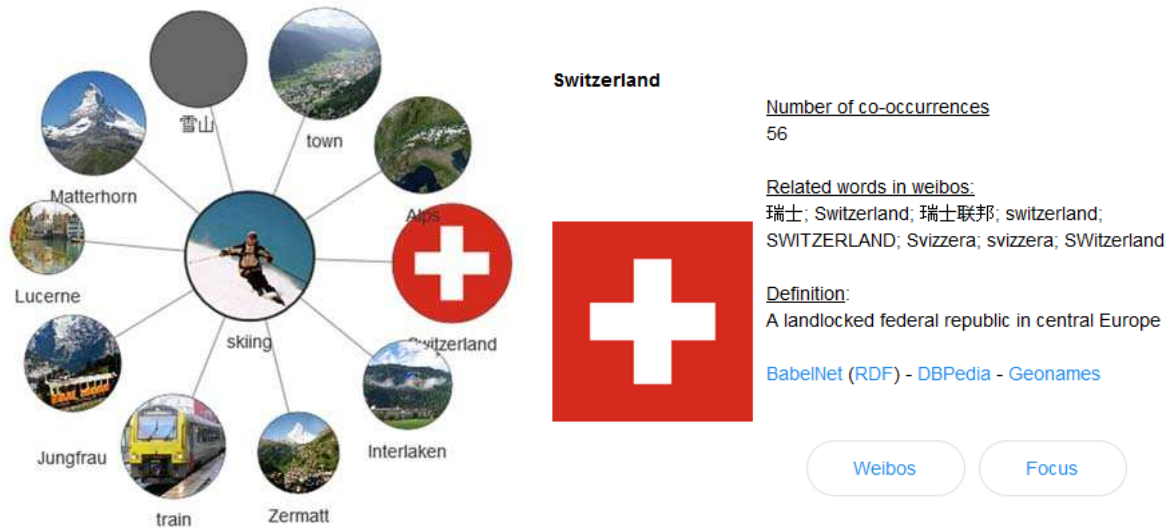
Fig. 7. Example of a semantic search

5. Conclusion and Implications

We adopted a linked data approach in this study to perform a large-scale analysis of the social media platform, Sina Weibo, to provide a better understanding of Chinese tourists' behaviors in, and attitudes about, Switzerland. We addressed three research questions: who are these tourists and where do they travel in Switzerland? What are the new trends among Chinese tourists in Switzerland? And, more importantly, how does a semantic-based methodology help us achieve a better understanding of travelers' experiences with Swiss tourism? We presented a semantic technology-based methodology to process users' interpretations in the form of unstructured text, which can help ensure that vocabularies of natural language expressions are mapped correctly to the existing ontologies.

This study makes a significant contribution to both the scientific community and the tourism industry. First, our findings filled an important gap in the academic literature on Chinese tourist behavior in Switzerland based on social media platforms. To date, little has been reported in the context of tourism-related UGC on Sina Weibo. Thus, this research serves as a first step in increasing our understanding of this phenomenon. While the majority of Chinese tourists were from economically developed regions, including Beijing, Shanghai, and Guangdong, the number of tourists from those cities has declined over the past three years. Our content analysis showed not only the tourists' experiences with products and services, and their trends between 2013 and 2015, but also identified the similarities and differences between first time and multiple time visitors. Chinese people appreciate Switzerland's natural beauty and luxury consumer goods greatly, as well as the fact that it is the headquarters of many organizations. From a methodological point of view, we highlighted the ways in which the Semantic Web and linked data technologies have the potential to help improve NLP through rules and queries. Our proposed methodology facilitated the integration of tourism information that originates from social media sources, and represents notions and conceptual relationships in the tourism domain. A set of ontologies was applied to generate an annotation thesaurus among multilingual texts automatically. Moreover, semantic and NLP technologies were used to extract a knowledge base that was presented as a set of entities and their relationships

in an ontological triple store. In particular, the use of graph models underlying RDF and SPARQL extensions enabled us to extract and parameterize the classic features of social media analysis from these representations directly and efficiently. Furthermore, the semantic search engine supports multiple languages and provides users with a more intelligent search than does a full-text search. Thus, it can serve as a feasible and practical approach for big data social media analysis in other domains.

From the perspective of the tourism industry, this research served to illustrate Chinese tourists' decision-making, and identified the main factors that influence travel decisions, and profiled growing areas and market trends. Our findings showed the primary characteristics of Chinese tourists in Switzerland based on the most popular Chinese SNS—Sina Weibo—and offered insightful implications about destination marketing and visitors' behaviors. Specifically, these can be summarized as follows: (1) Know visitors before they arrive. Our analytical results highlighted the places and activities that Chinese visitors in Switzerland find most interesting. It provided significant and useful information for destination marketing organizations to tailor marketing messages in the right way and in the right places. Chinese visitors often posted messages or photos when they arrived at an airport, which offers a good opportunity for tourism marketers to connect with them, and guide them according to their needs; (2) Learn about and plan for dissatisfied tourists. This study identified both positive and negative travel experiences and statistics from Chinese tourists in different places, hotels, restaurants, and activities in Switzerland. By attending to these messages, and especially to complaints, marketers can transform an adverse situation into a winning customer experience. Some will even serve as strong advocates in social media for destination marketing. Moreover, our approach provides a new direction to connect and analyze big data, and more importantly, convert it into business insights in real time. This will help decision-makers understand complex information intuitively, know the current operating environment, and the way in which the future may unfold. Finally, (3) use targeted marketing strategies and technology. The semantic-based social search engine has become one of the fastest methods for destination managers to find, recognize, link, and manage the resources of Chinese tourists' travel experiences in Switzerland. Such an advanced technological tool makes tourism-related products and services much more visible and distinguishable. Further, it allows targeted strategies to be developed for the right tourists, and increases the value of destination marketing significantly.

Swiss tourism agencies can also profit from our results, as they can obtain valuable insights about tourism marketing strategies, especially ways in which to communicate with individual tourists to attract independent travelers. For example, there is some evidence to suggest that Weibo messages about travel experiences are biased towards females, especially young women between the ages of 16–35 years. This demographic profile has implications for organizations who want to use Weibo as a marketing tool for specific products and services. More generally, our findings provided more reliable evidence than do certain technical reports available from officials, which are based purely on statistics and numbers.

Acknowledgements

References

Auer S, Bizer C, Kobilarov G, et al (2007) DBpedia: A Nucleus for a Web of Open Data. In: Proceedings of the 6th International The Semantic Web and 2Nd Asian Conference on Asian Semantic Web Conference. Springer-Verlag, Berlin, Heidelberg, pp 722–735

Bai S, Hao B, Li A, et al (2013) Predicting Big Five Personality Traits of Microblog Users. In: 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). pp 501–508

Bontcheva K, Cunningham H (2011) Semantic annotations and retrieval: Manual, semiautomatic, and automatic generation. In Handbook of semantic web technologies. pp. 77-116. Springer Berlin Heidelberg.

Bontcheva K, Rout D (2014) Making sense of social media streams through semantics: a survey. Semantic Web, 5:373-403

Burgess S, Sellitto C, Cox C, Buultjens J (2009) User-generated content (UGC) in tourism: benefits and concerns of online consumers. In: Information systems in a globalising world: challenges, ethics and practices: Proceedings of the 17th European Conference on Information Systems. Verona, Italy

Cheng M, Edwards D (2015). Social media in tourism: a visual analytic approach. Current Issues in Tourism 18:1080-1087. doi: 10.1080/13683500.2015.1036009

China Internet Watch (2015) Sina Weibo revenue and statistics. http://www.businessofapps.com/sina-weibo-revenue-and-statistics/ Accessed 09 March 2016

Chua T-S, Luan H, Sun M, Yang S (2012) NExT: NUS-Tsinghua Center for Extreme Search of User-Generated Content. IEEE Multimed 19:81–87

Cox C, Burgess S, Sellitto C, Buultjens J (2009) The Role of User-Generated Content in Tourists' Travel Planning Behavior. J Hosp Mark Manag 18:743–764. doi: 10.1080/19368620903235753

Djuana E, Xu Y, Li Y (2011) Constructing tag ontology from folksonomy based on WordNet. In: Kommers P, Zhang J, Issa T, Isaías P (eds) Proceedings of the IADIS International Conference on Internet Technologies and Society 2011. International Association for Development of the Information Society (IADIS), The East China Normal University, Shanghai

Doyd D, Crawford K (2012) Critical Questions for Big Data. Inf Commun Soc 15:662–679. doi: 10.1080/1369118X.2012.678878

Forrester (2012) Forrsights Business Decision-Makers Survey Q4 2012. https://www.forrester.com/Forrsights+Business+DecisionMakers+Survey+Q4+2012/-/E-SUS1851 Accessed 09 March 2016

Francalanci C, Hussain A (2016). Discovering social influencers with network visualization: evidence from the tourism domain. Information Technology & Tourism 16:103-125. doi:10.1007/s40558-015-0030-3

Fuchs, M, Höpken W, Lexhagen M (2014). Big data analytics for knowledge generation in tourism destinations– A case from Sweden. Journal of Destination Marketing & Management 3:198-209. doi: 10.1016/j.jdmm.2014.08.002

Gabrilovich E, Markovitch S (2007) Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 1606–1611

Garcia-Crespo A, Lopez-Cuadrado JL, Colomo-Palacios R, Gonzalez-Carrasco I, Ruiz-Mezcua B (2011). Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain. Expert systems with applications, 38:13310-13319. doi: 10.1016/j.eswa.2011.04.152

Guan W, Gao H, Yang M, et al (2014) Analyzing user behavior of the micro-blogging website Sina Weibo during hot social events. Phys Stat Mech Its Appl 395:340–351. doi: 10.1016/j.physa.2013.09.059

Hepp M (2010). HyperTwitter: collaborative knowledge engineering via twitter messages. In International Conference on Knowledge Engineering and Knowledge Management. pp. 451-461. Springer Berlin Heidelberg.

Hu T, Kalbaska N, Cantoni L (2013) A Picturesque Online Representation of Switzerland: A Research Conducted on Chinese Social network Weibo. http://www.sinoptic.ch/textes/recherche/2013/201304_USI_swiss-images-cn.pdf Accessed 09 March 2016

Hwang HS, Choi EK (2016). Exploring gender differences in motivations for using Sina Weibo. KSII Transactions on Internet and Information Systems 10:1429-1441.

Jacobsen JKS, Munar AM (2012) Tourist information search and destination choice in a digital age. Tour Manag Perspect 1:39–47. doi: 10.1016/j.tmp.2011.12.005

Jansen BJ, Zhang M, Sobel K, Chowdury A (2009) Twitter power: Tweets as electronic word of mouth. J Am Soc Inf Sci Technol 60:2169–2188. doi: 10.1002/asi.21149

Kiryakov A, Popov B, Terziev I, Manov D, Ognyanoff D (2004). Semantic annotation, indexing, and retrieval. Web Semantics: Science, Services and Agents on the World Wide Web, 2:49-79

Kleinen A, Scherp A, Staab S (2014). Interactive faceted search and exploration of open social media data on a touchscreen mobile phone. Multimedia tools and applications, 71:39-60. doi:10.1007/s11042-013-1366-3

Kobilaro, G, Scott T, Raimond Y, Oliver S, Sizemore C, Smethurst M, Bizer C, Lee R (2009). Media meets semantic web–how the bbc uses dbpedia and linked data to make connections. In The semantic web: research and applications. Springer Berlin Heidelberg. pp 723-737

Lazer D, Pentland A, Adamic L, et al (2009) Social science. Computational social science. Science 323:721–723. doi: 10.1126/science.1167742

Li Y, Bontcheva K (2007). Hierarchical, perceptron-like learning for ontology-based information extraction. In Proceedings of the 16th international conference on World Wide Web. pp. 777-786. ACM

Litvin SW, Goldsmith RE, Pan B (2008) Electronic word-of-mouth in hospitality and tourism management. Tour Manag 29:458–468. doi: 10.1016/j.tourman.2007.05.011

Li Y, Gao H, Yang M, et al (2015) What are Chinese talking about in hot weibos? Phys Stat Mech Its Appl 419:546–557. doi: 10.1016/j.physa.2014.10.043

MakeAWebSiteHub (2015) 2015 Chinese Social Media Statistics And Trends Infographic. http://makeawebsitehub.com/chinese-social-media-statistics/ Accessed 09 March 2016

Manyika J, Chui M, Brown B, et al (2011) Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company. http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation. Accessed 8 Mar 2016

Marr B (2015) Why only one of the 5 Vs of big data really matters. http://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters. Accessed 01 Juin 2016

Miguéns J, Baggio R, Costa C (2008). Social media and tourism destinations: TripAdvisor case study. Advances in Tourism Research 26:26-28. Doi: 10.6092/1970-9870/2501

Miller GA (1995) WordNet: A Lexical Database for English. Commun ACM 38:39–41. doi: 10.1145/219717.219748

Moro A, Cecconi F, Navigli R (2014a) Multilingual Word Sense Disambiguation and Entity Linking for Everybody. In: Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272. CEUR-WS.org, Aachen, Germany, Germany, pp 25–28

Moro A, Raganato A, Navigli R (2014b) Entity Linking meets Word Sense Disambiguation: A Unified Approach. Trans Assoc Comput Linguist 2:231–244

Muscanell NL, Guadagno RE (2012). Make new friends or keep the old: gender and personality differences in social networking use. Computers in Human Behavior 28:107-112

Navigli R, Ponzetto SP (2010) BabelNet: Building a Very Large Multilingual Semantic Network. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Stroudsburg, PA, USA, pp 216–225

Passant A, Laublet P (2008). Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data. In : Proceedings of the Linked Data on the Web Workshop (LDOW), Beijing, China.

Pérez J, Arenas M, Gutierrez C (2009) Semantics and Complexity of SPARQL. ACM Trans Database Syst 34:16:1–16:45. doi: 10.1145/1567274.1567278

Pirolli P (2009) Powers of 10: Modeling complex information seeking systems at multiple scales. IEEE Computer, 42:33-40

Pöschko J (2011) Exploring Twitter Hashtags. arXiv preprint arXiv:1111.6553

Shen H, Songm C, Li M, Jiang Q (2015).Shaping destination images through social networking sites: a case study of Singapore. Advances in Hospitality and Lesure 11:89-107

Simms A (2012) Online user-generated content for travel planning - different for different kinds of trips? E-Rev Tour Res 10:76–85

Switzerland Tourism (2013) Research Report China: Market Analysis and Insight. http://www.stnet.ch/files/?id=63853 Accessed 01 November 2014

Vicient C, Moreno A (2015) Unsupervised topic discovery in micro-blogging networks. Expert Syst Appl 42:6472–6485. doi: 10.1016/j.eswa.2015.04.014

Wick, M, Vatant, B (2012) The geonames geographical database. http://geonames. org Accessed 09 March 2016

Xiang Z, Gretzel U (2010). Role of social media in online travel information search. Tourism management 31: 179-188. doi: 10.1016/j.tourman.2009.02.016

Xiang Z, Schwartz Z, Gerdes JH, Uysal M (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? International Journal of Hospitality Management 44:120-130. doi: 10.1016/j.ijhm.2014.10.013

Yoo KH, Gretzel U (2010) Antecedents and Impacts of Trust in Travel-Related Consumer-Generated Media. Inf Technol Tour 12:139–152. doi: 10.3727/109830510X12887971002701

Yoo KH, Lee Y, Gretzel U, Fesenmaier DR (2009) Trust in Travel-related Consumer Generated Media. In: Höpken DW, Gretzel DU, Law DR (eds) Information and Communication Technologies in Tourism 2009. Springer Vienna, pp 49–59

Yu L, Asur S, Huberman BA (2011) What Trends in Chinese Social Media. arXiv preprint arXiv:1107.3522.

Zhang L, Pentina I (2012) Motivations and usage patterns of Weibo. Cyberpsychology Behavior and Social Networking 15:312–317. doi: 10.1089/cyber.2011.0615