

SOCIAL MEDIA MONITORING AND ANALYSIS OF A SWEDISH TOURISM DESTINATION

Christer Kuttainen
E-handelsdoktor
Piteå, Sweden
e-mail: christer@kuttainen.com

Maria Lexhagen
ETOUR, Mid Sweden University
Östersund, Sweden
e-mail: maria.lexhagen@miun.se

Matthias Fuchs
ETOUR, Mid Sweden University
Östersund, Sweden
e-mail: matthias.fuchs@miun.se

and

Wolfram Höpken
IAF, Hochschule Ravensburg-Weingarten
Weingarten, Germany
e-mail: wolfram.hoepken@hs-weingarten.de

ABSTRACT

Tourism destinations can benefit from systematically monitoring, collecting and analyzing user generated content (UGC). This exploratory study evaluates tools and methods for collecting UGC related to the leading Swedish ski destination of Åre, including its SMEs. Interviews with the largest individual stakeholder were conducted. Also, a workshop with 20 SMEs took place. Furthermore, text based UGC was collected from personal and travel related blogs, Twitter flows, travel review sites, special interest sites and a destination Facebook page between January and May 2011. Results indicate that both the DMO and SMEs would benefit from applying a basic social media monitoring toolbox.

Key Words: user generated content (UGC), online word-of-mouth, social media monitoring

INTRODUCTION

With the advent of web 2.0 applications the web has become an invaluable source of information for businesses and organizations (Cheng and Tseng, 2011). Since an important goal for DMOs (Destination management organizations) is to build and monitor the destination brand (Chekalina, Fuchs & Lexhagen, 2011) it is important to acknowledge that unofficial information sources, such as those available on the Internet, are gaining in popularity implying the need to manage a destinations online reputation from a holistic perspective (Inversini and Cantoni, 2011). However, taking the example of using travel blogs as a source of market information it has been suggested that this is difficult for DMOs' to do due to large variations in content, language barriers and cultural nuances in sentiment (Puhringer and Taylor 2008). Also, buying a ready-made technological solution for monitoring social media content is often not enough as a single method to use as input for branding strategies. In order to get results relevant for decision-making output from monitoring tools should be integrated with other business data, such as web statistics and guest satisfaction data (Birch and Karma, 2011). A further weakness of monitoring and using social media content as a source of marketing intelligence is also that unless the product or service in question is a "hot topic" the volume of online conversations is most likely not very extensive.

Nevertheless, the benefits would be that this type of information is perceived as unsolicited conversations about brands. Companies can use discussion forums both for passive market intelligence and for active market research. Marketers are advised to engage in specialized, relevant online communities. Here marketers can provide information, correct incorrect information and collect community member experiences. It

is, however, important to adhere to the rules of the specific community, for example avoid direct commercial messages (Pitta and Fowler, 2005). Collecting and analyzing web content, in particular word-of-mouth in social media, is suggested for both academics and marketing researchers in order to access information about customers. Such content is arguably both *rich* and *easily available*, and thus suitable for company decision making. UGC, however, must be evaluated and filtered by the researcher to establish its accuracy and validity (Chang, 2011).

Deciding what information to collect, from what source, and with which methods is, however, a risk-laden choice for any business (Adams, Richey, Harvey and Hilton, 2010). Destinations can use simple web searches to get a glimpse of how their brand is being portrayed but the risk is that a large volume of search results will deter them from efforts to deduce valuable knowledge from this. There are multiple tools available for collecting, categorizing, and analyzing online conversations. However, academic literature on social media monitoring appears scarce. The research question being posed in this study is whether it is feasible for a tourism destination to systematically monitor, collect and analyze user generated content in order to get useful insights for destination development. If so, how can this process be described?

LITERATURE REVIEW

Based on the dynamics of online word-of-mouth new strategies are needed. From an informational perspective, Litvin, Goldsmith and Pan (2008) propose that procedures that allow hospitality and tourism marketers to harvest discussion and feedback created online, are needed. Tsiotsou and Ratten (2010) contend that more research on decision-making models based on the emerging role of user-generated-content (UGC) and social networking is needed, implying also a need for methods of monitoring and collection of UGC. Accordingly, one vein of research on social media monitoring deals with automation and machine-learning along with methods to extract sentiment and prediction opportunities (Bai, 2011; Li and Wu, 2010), semantic oriented content analysis (Walchofer, Fröschl, Dippelreiter, Pöttler and Werthner, 2009), and sentiment and relational extraction (Xu, Liao, Li and Song, 2011).

Tourism research on UGC has also been undertaken to analyze for instance how tourists communicate their experience with a tourism destination (De Ascaniis and Morasso, 2011), how tourists in forum posts (e.g. TripAdvisor.com) argue for others to visit a certain destination (Fedele, De Ascaniis and Cantoni, 2011), or the use of review platforms (e.g. HolidayCheck.com) for complaint management (Maurer and Schaich, 2011). Specifically O'Connor (2010) found that few hotels are actively managing their reputation on the TripAdvisor site. He therefore suggests that it is important to increase activity in both monitoring and managing this type of UGC and if not, then competitive advantage will be lost.

It has also been stated that increasing volumes of social media content allow for businesses to cost-effectively monitor and use this as a source of marketing intelligence (Decker and Trusov, 2010; Chang, 2011; Pan, MacLaurin and Crotts, 2007). Additionally, Dickinger, Költringer, and Körbitz (2011) propose a method to automatically collect thousands of posts from international travel communities and travel guides. They found that the results of text mining blogs reach similar results as traditional image studies but the benefits are that analyzing UGC is less intrusive. None of these studies, however, deals with the question of how a DMO can make the same kind of UGC monitoring, collection, and analysis as the researchers do. Research studies as mentioned above concentrate on one or a few sources of UGC. By contrast a DMO in real life would most likely encounter multi-faceted UGC both in terms of sources and characteristics. Reviewed research studies also lack in detail regarding the concrete UGC monitoring and collection process, including time and resources spent. However, such information would be highly interesting from a practitioners' perspective. One of few academic papers that was identified for this present study dealing with tools and methods for collecting UGC, is Laine and Frürwirth (2010). They found that social media monitoring tools are still in their infancy, with much room for improvement. Thus companies are advised to carefully choose and evaluate the benefits of these tools.

METHODOLOGY

This paper presents a study on tools and methods for collecting user generated content related to the leading Swedish ski destination of Åre. The applied research methodology can be classified as a case study using an exploratory research approach in three steps. First, interviews with the largest individual stakeholder (SkiStar Åre) were conducted in order to understand current use of social media marketing and monitoring. Furthermore, text based data was collected from personal and travel related blogs, Twitter flows, travel review sites, special interest sites and a company Facebook group. In addition, a workshop was held after the UGC data collection was complete. The workshop involved around 20 SME representatives in the Åre region.

Table 1
Social media monitoring and analysis tools used in the case study

SMMA tool	Comments
Trackur	Free version used. Web based. Indexing of blogs, video, Twitter (not Facebook, Google + or forums). Sentiment indicator. Keywords: "Åre, Skistar" (can not index letter "Å"). Used initially to get an overview of UGC, no results retained for analysis due to inability of indexing keyword Åre and no result export feature.
Social Mention	Free. Web based. Sentiment and other metrics. Keywords: "Åre, Skistar" (can not index letter "Å"). Used initially to get an overview of UGC, no results retained for analysis due to inability of indexing keyword "Åre", and unreliable result export feature.
Google Alerts	Free. Keywords: "Åre + Skistar + [20 SMEs in Åre], appr. 1,000 hits, 140 retained. Requires manual scanning of results in e-mail and manual export.
SM2 Alterian	Free version used (<1,000 results presented). Web based, Indexing of most sources, including Facebook. Dashboard, useful hit filtering and export features. Keywords: "Åre, Skistar". 100 hits retained.
Tweettronics	Version Pro (\$50/month) Scanning and capture of tweets. Web based. Keywords: "Åreskiresort, Skistar, + 12 industry specific keywords (can not index letter "Å") Appr. 12,000 tweets retained
Mozenda	Licensed tool (\$399 annual plan used). Web scraping software. Used for automatic collection of web content at chosen sources. Scraped part of the review site content. Requires programming skills to work flawlessly on all sites. Certain web sites block web scraping.

In step 2 social media monitoring and analysis tools (SMMA) were used to identify and scan content (Table 1). SMMA tools were chosen with the intent to evaluate them from a DMO as well as a SME perspective. For this reason our focus was on "low spec – low cost" tools. This is defined as tools that are free or near free in cost (Stevenson, 2010). A limited tool set was applied and evaluated as the data collection progressed. The idea was to mimic how a destination (or a SME) would possibly apply SMMA in real-life.

Data was located and collected using the tools Google Alerts, SM2 Alterian (Laine and Frühwirth, 2010) and Tweettronics for capturing Twitter content, as well as the web scraping software Mozenda (for most of the review sites and the special interest site). UGC was located based on both known sources (e.g Freeride and Skistar Facebook page and Twitter account) and hitherto unknown sources as presented by the SMMA tools' search results. To find new UGC sources primarily keywords "Åre" and "Skistar" were provided. Keywords were added in the form of names of all the member companies of the destination organization (Åre Destination AB) with a primary focus on tourism services. In addition around 15 industry related keywords were used, for example "skiing, mountain, hiking" in combination with "Sweden" or "Scandinavia". Data collection of UGC then targeted special interest sites, review sites, travel forums, other forums, personal blogs and Skistar's Facebook page.

The data collection period extended from January 1 to May 2, 2011, which corresponds to the winter season in Åre. Additional data (*review sites*) was collected for the period Sept – Dec 2010, which can be considered pre-winter season in Åre. Due to very large volumes of some of the search words, Twitter content was only collected for the period 11th of April until 3rd of May.

UGC from travel forums and other relatively comprehensive sources of data was extracted using the web scraping software Mozenda. The reason for using Mozenda is that it greatly reduces the effort of collecting data, compared to tedious manual procedures as for example in Carson & Schmalleger (2009). Sources with a relatively small number of relevant postings were manually extracted, essentially a copy-paste procedure.

Finally, UGC data files were created in an Excel file format and cleansed from irrelevant content. For example, the Google Alerts file contained around 1,000 entries, after cleansing it was reduced to 140 posts. A post was the smallest unit of analysis, usually consisting of a few sentences. UGC data files were imported into the software NVivo for coding and analysis. UGC data was roughly structured according to (Cao, Duan and Gan, 2011): 1) userID 2) review post time 3) review title 4) review content 5) votes/ratings (if evident) 6) other.

RESULTS

Results from step 1

As part of the case study approach, representatives from the largest company at the Åre destination, SkiStar, were interviewed (December 2010). The objective was to learn if social media monitoring and analysis is currently conducted, understand possible needs and motives for SMMA and also to get input on how to practically deploy a SMMA study for Åre - SkiStar. The company SkiStar is a public company, whose core business is alpine skiing. Skistar operates five destinations in Sweden and Norway, one of which is Åre. Company knowledge about the customer is available mainly in the booking and marketing departments. As for the use of SMMA tools Google Alerts is used on an ad-hoc basis by the person responsible for social media scanning and management of content on the company Facebook page and on Twitter. This person also manually looks for mentions of Åre and SkiStar on the web. SkiStar management would like to gain greater knowledge about the customer from different social media channels. For instance, what are the top three issues being discussed by Skistar's customers? The company can then include such issues in its product development. Furthermore, SkiStar would like to more clearly compile customer experiences and needs. One way put forward is to segment customers based on their use of social media, as opposed to traditional demographic profiling, according to the SkiStar representatives.

Results from step 2

Results are based on categorizing and evaluating multiple SMMA tools in relation to their compatibility with the needs of a DMO to build and monitor the destination brand.

Table 2
Sources of UGC and number of posts collected

Source	No. of posts
<i>Destination specific sites</i>	
Åre-Skistar Facebook page	1 234
Snöbloggen	8
<i>Special interest site</i>	
Freeride	291
<i>Review sites</i>	
Booking.com	162
Tripadvisor.com	41
Rejta (Hotel, Hostel, Conference)	16
Rejta (Eat out)	25
Rejta (Other)	13
Restaurangkartan.se	10
<i>Content by monitoring tools</i>	
Google Alerts	141
SM2 Alterian	102
<i>Total – web sites</i>	<i>2 043</i>
<i>Tweets</i>	
@Areskiresort	9
Areskiresort	39
@Skistar	3
Skistar	187
fjällen	475
fjällvandring	10
mountain biking	2 044
mountain hiking	404
ski holiday	86
Skidresa	9
skidsemester	6
Skiing	9 079
skiing scandinavia	4
skiing sweden	9
<i>Total - tweets</i>	<i>12 364</i>

Total	14 407
-------	--------

With the chosen set of SMMA tools different sources of UGC about Åre, the company SkiStar and other key words were detected. Around 2,000 relevant web site postings and 12,000 tweets were generated through a combination of using tools and manual collection efforts (Table 2). The use of multiple tools was deemed necessary to collect relevant UGC from different sources. For example, the tools differ in their results presentation of personal blogs during the same search period. No “high spec, high cost” (Stevenson 2010) SMMA solution was applied due to the prohibitive costs (both for researchers and SMEs), in the region of \$10,000.

Web scraping software, such as Mozenda, can greatly assist in collecting large quantities of UGC from known sources. Web scraping can also be scheduled and, thus, automate UGC collection processes. However, based on our experience with Mozenda, successful application of web scraping demands a great deal of programming skills. Some web sites can be scraped via the Mozenda interface while other web sites, more complicated in their structure, demands deep knowledge of web interface coding and the ability to modify web scraping procedures accordingly. There are also issues such as certain web sites block the use of automated data collection. For these reasons web scraping appears useful primarily on a destination level, where appropriate resources can be dedicated or the service out-sourced.

Thus, extensive manual efforts were required in all steps of the monitoring, collection, and analysis process. Examples include choosing and refining search keywords, setting up monitoring and collection procedures, transfer and convert data sets, as well as transforming content into usable information. Automatic sentiment analysis relies on a list of words which is compared with the occurrence of certain words in the collected data (Dickinger et al., 2011). No such sentiment analysis was made in this study. Instead, the researchers manually evaluated each post, to get a deeper, more holistic understanding of the actual sentiment of the UGC. Similarly, Fedele et al (2011) argue that analysis of UGC requires human intervention. The claim is that only humans can adequately determine the meaning of a text, and the intended message in a specific context. Based on this assumption the results of this study were obtained through a manual evaluation process of the type and quality of collected UGC.

UGC data collection covered the Åre-SkiStar winter season, including typical demand fluctuations. Thus SMMA provided an overall picture of UGC sources and typical content. This picture can be said to constitute a baseline for future comparison of UGC activity for the winter season.

In terms of obtained UGC volume and relevance in relation to collection efforts, it can be said that several of the used low spec – low cost SMMA tools are indeed very simple to use and results appear trustworthy and suitable for further analysis. To these tools belongs Google Alerts. Other tools raise questions about functionality and validity. Of the tools employed in this study, we stopped using Trackur and Social Mention at an early stage of the data collection. The main reasons were cumbersome user interface, including questions about the validity of the results, and no or badly functioning results export capability (see Table 1).

Results from step 3

A workshop was held with around 20 representatives from SMEs in the Åre region. The objective of the workshop was twofold: 1) extend the study scope to SMEs, which constitute all of the businesses at the destination except for SkiStar, 2) present the case of SMMA for Åre and get their feedback on the study results and tools used. These SMEs represented tourism, manufacturing, trade and other services. During the workshop preliminary study results were presented, among them the type and quality of UGC related to Åre, Skistar and also related to local SMEs, some of which were present at the workshop. It was demonstrated that through a simple Google Alerts scan of the company names of Åre SMEs, the sources, quantity and type of UGC can be relatively easily established. For one of the participating companies specifically relevant blogs were revealed through the SMMA tools.

The workshop participants also testified to the researchers that SMMA among SMEs is either not done at all or only superficially. Only a few of the participating companies employed SMMA. A handful had briefly used Google Alerts, but not in a systematic manner. One of the participants expressed the view that she had not even considered the possibility of actually listening to UGC. The main impression from the workshop was that Åre SMEs were positive to the idea and exhibited a willingness to start with or develop their use of SMMA. Much discussion also revolved around how to start and maintain a Facebook fan (company) page. This discussion circled around the issue of acquiring customer knowledge from fan related UGC, as opposed to

discover new sources for UGC. A few workshop participants could also confirm that an active Facebook presence results in business.

CONCLUSIONS AND IMPLICATIONS

A first conclusion is that DMOs and SMEs in particular, due to limited resources, are recommended to start using the most-easy-to-use tools in order to establish a working routine for collection and analysis. Low cost, low spec SMMA tools appear functional enough to give the SME or tourism destination manager a “good enough” overview of UGC sources and their content. In this process the normal or baseline UGC situation can be established.

The argument for beginning SMMA in the low cost, low spec segment is the case in particular because several highly relevant forums, review sites and special interest sites are either well-known by the destination or requires only minimal monitoring to locate. It is, however, also evident that within the low cost, low spec category all free tools might not yield enough value. For this reason, a low cost solution (as opposed to free) is likely needed in the SMMA tool box. For instance, a result export feature is vital in order to be able to distribute and more effectively analyze results across organizational departments. A basic SMMA tool box for an SME or a tourism destination could contain 1) Google Alerts (because of its ease-of-use, and apparently valid results), 2) a low cost SM tool of choice, 3) a Twitter capturing tool, (if this source is deemed relevant, and the capture is not included in the low cost SMMA tool).

Google Alerts is recommended because of its extreme simplicity to get started. The Google Alerts e-mail snippets, however, are presently not possible to export to a database for further distribution in the organization. Labor intensive work is required, essentially cut-and-paste into a spreadsheet, thus reducing the value of the Alerts in terms of organizational decision-making. For this reason it is advisable to use another tool with good data export features. Furthermore, dedicated Facebook and Google + monitoring features can be considered, depending on where relevant sources and content are located. The landscape of SMMA tools is quickly changing, with new tools or services being introduced in a rapid fashion, and existent tools getting new features and cost alternatives (from limited free versions to full-feature, costly premium versions). For this reason an organization should carefully evaluate SMMA tool, as Laine and Frürwirth (2011) highlight.

Social media monitoring should be automated if possible. For example, the Google search engine can be used to manually find recent blog entries. The same results appear to be presented in Google Alerts. However, it is unlikely that a DMO or a SME owner-manager would regularly “google” company related key words. Google Alerts, on the other hand automatically presents the result in the company email, thus raising awareness of relevant UGC.

The use of SMMA can reveal specialized online communities relevant for a tourism destination. While the most popular communities are familiar to destination management, for example the Freeride skiing forum, via SMMA tools management can discover other relevant forums, and then possibly engage to provide information, if appropriate, or simply passively follow activity to get ideas for product development (Pitta and Fowler 1995).

The researchers in this study would like to acknowledge the support from *Stiftelsen för Kunskaps och Kompetensutveckling* (KK-stiftelsen) for the funding of the project *Engineering the Knowledge Destination*.

REFERENCES

- Adams, F. G., Richey, R. G. Jr., Harvey, M. G., & Hilton, C. B. (2010). Information, intelligence, and resource advantage: a multi-market multi-theoretic call for research. *International Journal of Transitions and Innovation Systems*, 1(1): 4-24.
- Bai, X. (2011). Predicting consumer sentiments from online text. *Decision Support Systems*, 50: 732-742.
- Birch, D., & Karma, B. (2011). Case studies in social media listening and optimization. *e-Review of Tourism Research*, 9(1): 9-11.
- Cao, Q., Duan, W. & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2): 511-521.
- Carson, D. & Schmallegger, D. (2009). Destination Image Projection on Consumer-Generated Content Websites: A Case Study of the Flinders Ranges. *Information Technology & Tourism*, 11(2): 111-127.
- Chang, J. (2011). Conceptualizing the value of the Web content in marketing research. *Marketing intelligence & Planning*, 29(7): 687-696.

- Chekalina, T., Fuchs, M. & Lexhagen, M. (2011) Determinants of the co-created destination experience: an empirical validation from Sweden, *Advances in Tourism Marketing conference – Transforming Experiences Tourism Marketing from both Sides of the Counter*, 6th – 8th of September 2011, Maribor, Slovenia.
- Chen, C. C., & Tseng, Y. D. (2011). Quality evaluation of product reviews using an information quality framework. *Decision Support Systems*, 50(4): 755-768.
- De Ascaniis, S. & Morasso, S.G. (2011). When tourists give their reasons on the Web: The argumentative significance of tourism related UGC. In R. Law, M. Fuchs, F. Ricci (Eds.), *Information and Communication Technologies in Tourism 2011* (pp. 125-137), Proceedings of the International Conference in Innsbruck, Austria, January: 26-28.
- Decker, R. & Trusov, M. (2010). Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4): 293-307.
- Dickinger, A., Költringer, C. and Körbitz, W. (2011). Comparing Online Destination Image with Conventional Image Measurement - The Case of Tallinn. In R. Law, F. Ricci and M. Fuchs (Eds.), *Information and Communication Technologies in Tourism 2011* (pp. 165-177). Proceedings of the International Conference in Innsbruck, Austria, January: 26-28.
- Fedele, S., De Ascaniis, S. & Cantoni, L. (2011). Destination Marketing and User's Appraisal: Looking for the reasons why tourists like a destination. In R. Law, F. Ricci and M. Fuchs (Eds.), *Information and Communication Technologies in Tourism 2011* (pp. 151-163). Proceedings of the International Conference in Innsbruck, Austria, January: 26-28.
- Inversini, A. & Cantoni, L. (2011). Towards online content classification in understanding tourism destination information competition and reputation. *International Journal of Internet Marketing and Advertising*, 6(3): 282-299.
- Laine, M. & Frühwirth, C. (2010). Monitoring Social Media: Tools, Characteristics and Implications. *Business Information Processing*, 51(2): 193-198.
- Li, N. & Wu, D. D. (2010). Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems*, 48(2): 354-368.
- Litvin, S. W., Goldsmith R. E. & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3): 458-468.
- Maurer, C. & Schaich, S. (2011). Online Customer Reviews Used as Complaint Management Tool. In R. Law, F. Ricci and M. Fuchs (Eds.), *Information and Communication Technologies in Tourism 2011* (pp. 499-511). Proceedings of the International Conference in Innsbruck, Austria, January 26-28.
- O'Connor, P. (2010). Managing a hotel's image on TripAdvisor. *Journal of Hospitality Marketing & Management*, 19(7): 754-772.
- Pan, B., MacLaurin, T. & Crotts, J.C. (2007). Travel blogs and the implications for destination marketing. *Journal of Travel Research*, 46(35): 35-45.
- Pitta, D. A. & Fowler, D. (2005). Internet community forums: an untapped resource for consumer marketers. *Journal of Consumer Marketing*, 22(5): 265-274.
- Puhringer, S. & Taylor, A. (2008). A practitioner's report on blogs as a potential source of destination marketing intelligence. *Journal of Vacation Marketing*, 14(2): 177-187.
- Stevenson, A. (2010), Social Media Monitoring Tools Demystified: Analysing Sources. <http://energise2-0.com/2010/09/30/social-media-monitoring-tools-demystified-analysing-sources/> [Accessed the 11th of January 2012, 14:55]
- Tsiotsou, R. & Ratten, V. (2010). Future research directions in tourism marketing. *Marketing Intelligence & Planning*, 28(4): 533-544.
- Xu, K., Liao, S. S., Li, J., & Song, Y. (2011). Mining comparative opinions from customer reviews for competitive intelligence. *Decision Support Systems*, 50(4): 743-754.