Exploring User Generated Data Visualization in the Accommodation Sector

Elizabeth Carvalho, Marcirio Silveira Chaves
Business and Information Technology Research Center (BITREC),
Universidade Atlântica, Fábrica da Pólvora de Barcarena, Barcarena, Portugal
{ecarvalho|marcirioc}@uatlantica.pt

Abstract

An important part of our information-gathering behavior has always been to find out what other people think. People now actively use information technologies to seek out and understand the opinions of others. This is due to the availability and popularity of opinion-rich resources such as online review sites and personal blogs. The accommodation sector is one of those where the guest opinion dictates its future. This sector needs to adapt and update the services they offer to their guests in order to stay competitive.

In this paper we outline the potential use of tree oriented visual techniques to map 1500 online reviews collected from 50 small and medium hotels (SMH). Our goal is to deliver visually the results of these reviews as straightforward and intuitively as possible to the accommodation managers, enabling them use it to support their decision making.

Keywords--- **Online Reviews, Small and Medium Hotels, Accommodation Sector, Information Visualization.**

1. Introduction

"What other people think" has always been an important piece of information during a decision-making process. The Web has made possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. They are making their opinions available to strangers via the Internet through online reviews, blogs and social networks.

This information is highly valuable to "tune" services and even "tailor" new ones that are offered to the public. In the tourism and hospitality industries, accommodation managers have a special role, since they are the responsible for the closest contact to the voyager. To have an efficient management of their services means to be proactive and know as soon as possible what is the prospects' or guests' feeling towards the services offered by the accommodation.

This work demonstrates the application of information visualization techniques to user generated content (UGC), specifically to the accommodation sector. We used online reviews collected from guests of small and medium hotels (SMH). We cross-reference this data to find patterns that could give a better insight and support an effective decision-making. To classify and analyze the data, we developed a methodology based on concepts of a domain-specific ontology — Hontology, described in detail in [23].

At this stage of our work, we want to elicit formal requirements for visual design of the data. To achieve this, we produced some graphs to work as our test bed. The result from this stage will help us to redesign and define the visual output to be delivered by a tool that is under development in the scope of the framework proposed in [26].

We want to know better the following aspects:
- What is the visual model of the potential end-user?
- How should we properly map and render:
  o the most valued accommodation features?
  o the perception of the quality offered by the hotel?
  o the correlation between the guest's profile and the mostly relevant features?
  o the intensity of the positivity or negativity of the features?
- Does the use of advanced visual techniques (such as tree oriented) to map the results will help the accommodation managers to have a better insight of the data?

This paper is organized in five sections. Section 2 gives a brief background on the use of visualization techniques to map blog data and opinion surveys, and references related work. Section 3 presents the methodology applied to collect and treat the data. Section 4 introduces significant aspects of the visual design of data and describes how it was done. Section 5 concludes this study also giving an outlook on important future work.

2. Related Work

Recently, much effort has gone into automatic opinion mining, making it possible to obtain customer opinions from a large amount of free review text. The effective visualization of it has also been a primary goal.

Wu et al. [1] presented OpinionSeer, an interactive visualization system that could visually analyze a large collection of online hotel customer reviews. The system is built on a new visualization-centric opinion mining technique that considers uncertainty for faithfully modeling and analyzing customer opinions. A new visual representation was developed to convey customer opinions by augmenting well-established scatterplots and radial visualization.

Weaver et al. [2] implemented an interactive visual tool for exploring the visitation patterns of guests at two hotels in central Pennsylvania from 1894 to 1900. They implemented it as a coordinated multiple view visualization in Improvise. Numerous discoveries have driven additional data collection from archival newspaper and census sources, as well as plans to enhance analysis of spatial patterns using historic weather records and railroad schedules. Distributed online evaluations of usability and usefulness have resulted in feature and design recommendations that were incorporated into the tool.

Sung et al. [3] adopted a proposed heuristic n-phrase rule to identify the polarity customer's opinions and provide position maps to visualize the pros and cons of respective hotels. They concluded that some inherent limitations such as the weakness in blogs itself such as free format of text and the difficulty to generalize findings due to the small size of bloggers, majority are young people. They illustrated the whole opinion mining of hotel customer generated contents in Chinese-language weblogs and proposed technique for the automatic feature-opinion extraction as well as for the visualization of the detected opinions polarity on the different features in the accommodation sector.

Gregory et al. [4] described a technical approach for analyzing the content of blog data using a visual analytic tool, IN-SPIRE, developed by Pacific Northwest National Laboratory. They presented a methodology for blog analysis using a mature document visualization tool. With this tool, users can harvest blogs (datasets can be static or dynamic, updating with real time information), view them by thematic content, isolate key words of interest, run queries, visualize changes in content over time, or isolate bloggers of interest.

Bross, Schilf and Meinel [5] implemented POSTCONNECT — an Adobe-Flex application with the aid of Flare (http://flare.prefuse.org). They wanted to provide a contemporary and innovative tool for the improved exploration of blog archive. POSTCONNECT is an interactive and well-arranged visualization tool to powerfully explore and browse standard blog systems archives that is the synthesis and application of existing visualization and interaction techniques to the new domain of weblogs. Figure 1 shows the development phases of the visualization.
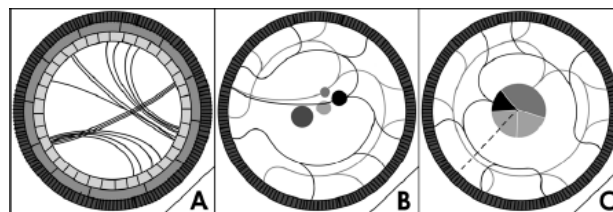
Figure 1 Development phases of visualization

3. Data and Methodology

Our data was obtained from an exploratory study conducted by [22], according to this methodology:

First phase: They gathered the names of all hotels in the Lisbon region registered in the National Registry of Tourism (http://bit.ly/vZy8rc) up to March 31, 2011 according to the following criteria: i) the hotel must be independent (i.e. not belong to a chain of hotels); ii) it must have fewer than 120 rooms; iii) it must have at least 30 reviews available for 2010 on Booking.com and Tripadvisor online pages, with text, in Portuguese, English or Spanish (for hotels that did not have enough comments for 2010, they supplemented them with reviews from the beginning of 2011). In cases where there were more than 30 reviews, we selected the comments with more description (i.e. the 30 longest comments). Due to fact that the number of hotels present in the National Register of Tourism that met all requirements fell short of the desired number (50 Hotels), they also used the list of housing program DiscoverPortugal (http://bit.ly/vrxgU3). This list comes from the site of Tourism of Portugal from April 15, 2011. They selected the first seven hotels in the city of Lisbon that met the criteria described above.

Phase two: They loaded 1500 reviews to a spreadsheet, and subsequently analysed them using the following parameters:
- Polarity: Identified taking into account six categories: positive, negative, mixed, neutral, irrelevant, or uncertain, as proposed by [24];
- Relevance of Concept of the Ontology (CO): They chose the segments of the reviews containing relevant CO (maximum three per review);
- CO: They identified the CO present in each segment selected;
- Strength of the Polarity (for each CO, polarity was identified in the following categories: very negative, negative, neutral, positive, or very positive. The strengths of the polarity were classified according to the approach of [25]. In this approach, the classification is chosen by the association of adjectives, adverbs of positive or negative connotation for the chosen parameter (e.g. clean, dirty, nice, nasty, comfortable and uncomfortable). For the classification of the strength very positive and very negative, the criterion used was the presence of adjectives or expressions of extreme tilt (e.g. stainless, dirty, good, bad, very, and could not have done more to);

- Qualifier of each CO, i.e. the term attributable to a positive or negative use of the CO.

In addition to classifying the reviews on the basis of the ontology, they identified what type of guest made the review. In this case, they used the categories provided by Booking.com and TripAdvisor sites. Other attributes that were evaluated were the number of stars of the hotel and the rating given by each user to the hotel.

Third phase: They analysed the sample data through Excel pivot table function. The data were analysed by means of the intersection of several variables. The purpose of this stage of the study was to identify possible patterns of market preference, which could support the decision-making of the SMH managers.

In this paper, we extend this methodology through the application of visualization methods according to specific goals of an accommodation manager.

4. Visualization Design

Creating visualization requires a number of nuanced judgments. One must determine which questions to ask, identify the appropriate data, and select effective visual encodings to map data values to graphical features such as position, size, shape, and color [6]. The challenge is that for any given data set the number of visual encodings—and thus the space of possible visualization designs—is extremely large.

There are a number of taxonomical studies on interaction in visualization, e.g., [7, 8, 9], and taxonomy proposals for specific classes of techniques and applications, e.g., [10, 11]. Brodlie proposed a notation for symbolic labeling visualization methods [12]. Card and Makinlay proposed a descriptive structure for visualization [13]. Duke et al. presented an argument to bring taxonomy and ontology together [14]. Other important examples are given by Bertin [15, 16], Cleveland [17] and Wilkinson [18].

The fundamental substrate of visualizations is spatial position. Marks such as points, lines, or area-covering elements can be placed on this substrate. These marks can carry additional information independent of their spatial position, such as size, greyscale luminance (brightness) value, surface texture density, color hue, color saturation, curvature, angle, and shape. The literature contains many different names for these kinds of visual encodings: retinal variables, retinal attribute, elementary graphical perception tasks, perceptual tasks, perceptual dimensions, perceptual channels, display channels, display dimensions, and so on. Section 4.1 describes the steps and what aspects were considered relevant to design the visualization of the data.

4.1. Design phases

The data collected is in tabular format. It contains fields like the hotel name, the number of stars and the total number of rooms. Although all the collected information is significant and interesting to be analysed, we selected only some fields to build the first group of visualizations. These are going to be used as the first rough sketch of the visual encodings to be tested against potential end-users. According the input that we will receive from them, our visualization design is going to evolve to more customized and enhanced versions.

To start the visualization design process we identified what relations in the data should be visually encoded and what were their goals. Table 1 summarizes the main conclusions.

Relation	Data	Goal
CO-P-R	Concept of the Ontology	To properly visually represent what are the features in the hotels that are mostly observed and relevant to the guests.
	Polarity	
	Rating	
CO-P-PS	CO	To infer what is the guest's perception of the quality offered by the hotels.
	Polarity	
	Polarity Strength	
G-CO-P	Guest	To spot the correlation between the guests's profile and the mostly relevant features.
	CO	
	Polarity	
CO-Q-PS	CO	To identify the intensity of the positive or negative use of a CO.
	Qualifier	
	Polarity Strength	

Table 1 Data and goals of the visualization

Data	Category	Type
CO	dimension	Nominal
Count of CO	measure	Quantitative
Polarity	dimension	Nominal and ordinal
Count of polarity	measure	Quantitative
Polarity strength	dimension	Nominal and ordinal
Count of polarity strength	measure	Nominal
Rating	dimension	Quantitative and ordinal
Count of rating	measure	Quantitative
Average of rating	measure	Quantitative
Guest	dimension	Nominal
Count of guests	measure	Quantitative
Qualifier	dimension	Nominal and ordinal
Count of guests	measure	Quantitative

Table 2 Data classification

The next step was to classify the data, in order to identify more straightforward what visual techniques would fit better to the data. Because the gathered data are discrete variables describing it (defines a domain), it was counted and averaged in order to derive its main measures. Table 2 presents the results of this step.

Finally, taking into account the guidelines given by the visual encoding taxonomies, we used the color, size and position as the main visual encodings. Additionally, we used text, legends, symbols and zooming to improve readability. Because the data describes a hierarchic relationship, we focused on the usage of tree oriented visualizations to map this – bubble, radial, treemap and force-direct, although we also produced plot graphs for comparison. To develop these rough sketches, we used off-the-shelf visualization tools: Tableau [19], Tulip [20] and Treemap [21]. Tableau Desktop is based on technology from Stanford University that lets drag & drop to analyze data. Tulip is an information visualization framework dedicated to the analysis and visualization of relational data. Treemap enables users to compare nodes and sub-trees even at varying depth in the tree, and help them spot patterns and exceptions. These tools offer interesting visual mapping solutions that allow a fast and flexible rendering of the data. Figures 1, 2 and 3 illustrate some of the visual results that were achieved.

Figure 1 presents two bubble tree graphs to depict the relation CO-P-PS. The color was used to map the polarity and the polarity strength values on the CO. The size was used to map the frequency that the CO is mentioned in the reviews. Besides this, symbols were applied in each node to map zones of influence of the rating or the polarity when the user zoom the image. Labels were also used to identify the CO on each bubble.

Figure 2 illustrates the mapping of the relation G-CO-P. We used a squarified treemap using the color, the size and the hierarchical ordering of the data to give different views on this relation. We also applied a slice and dice visual arrangement but the distribution of the data caused some parts of the tree to become a very thin slice, thus not giving a clear cue.

Finally, Figure 3 shows the visual mapping of the relation CO-P-R. Because we wanted to include a more

standard graph representation, we produced these plots. The thickness of the polyline, the size of the symbols, the position and the color are used to depict how the rating and the polarity impact on the CO.

4.2. Visualization Evaluation

Although the Information Visualization field is not new, it still suffers from a lack of methodologies to evaluate the results that it produces.

Catherine Pleasant [27] presents the challenges of the evaluation of the visualization and suggests three possible first steps to improve information visualization evaluation and facilitate adoption: the development of repositories of data and tasks, the gathering of case studies and success stories, and the strengthening of the role of toolkits.

Chen and Czerwinski [28] provide a timely and unifying forum for researchers and practitioners to tackle some of the fundamental and practical issues concerning empirical evaluation of information visualizations.

Most recently, Lam et al. [29] proposed seven guiding scenarios to evaluate information visualization. These scenarios were derived through an extensive literature review of over 800 visualization publications and are described through their goals, the types of questions they embody and illustrated through example studies.

Because the goal of information visualization evaluations in our research is to work towards understanding the work, analysis, or information processing practices by a group of hotel managers with or without software in use, we will adopt the first scenario proposed by Lam et al. The evaluations in the EWP - Evaluating Environments and Work Practices group elicit formal requirements for design, which fits perfectly our goal at this stage.
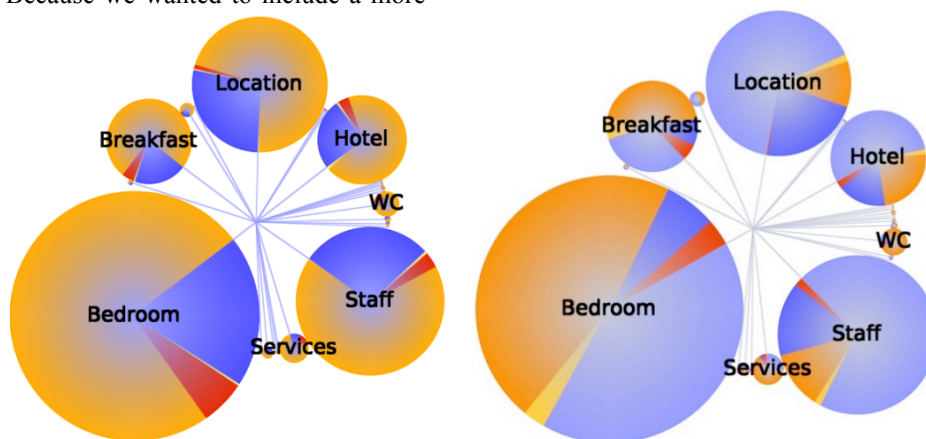
Figure1 Result of the application of Bubble Tree visualization of the relation among concepts of the ontology, polarity (left) and polarity strength (right).
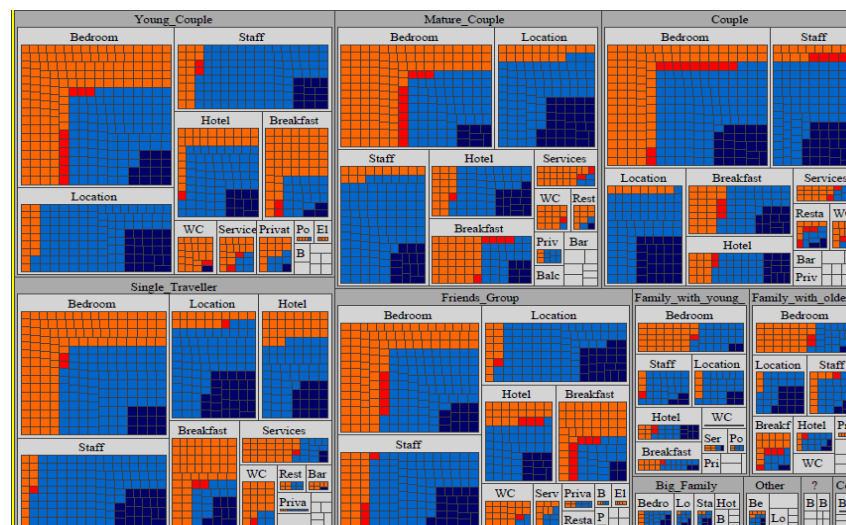
Figure 2 Results using Treemap visualization of the relation among guests, concepts of the ontology and polarity.
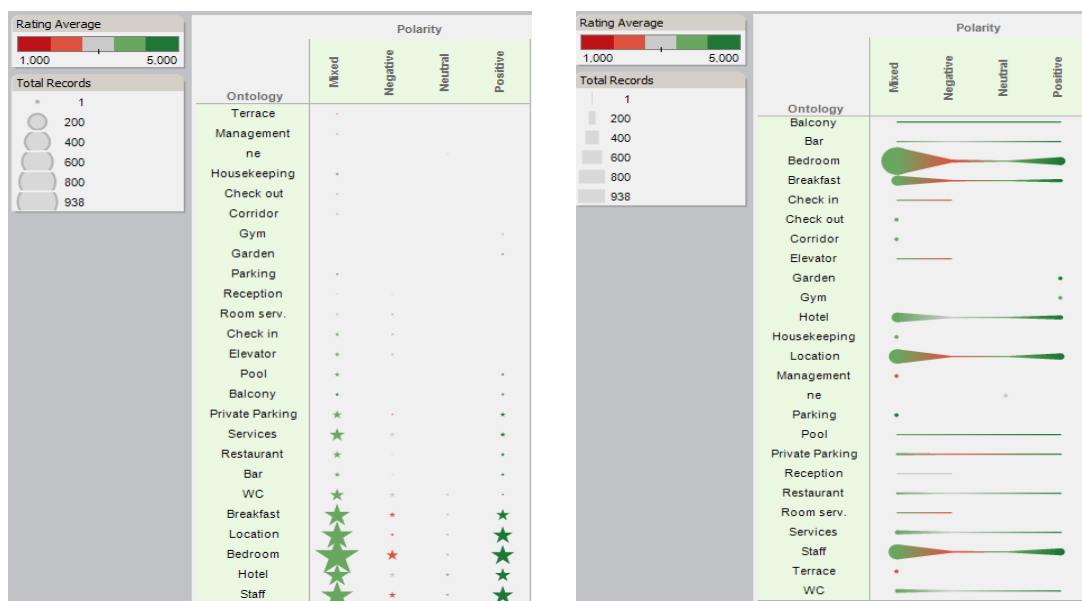

Figure 3 Results using Plot visualization of the relation among concepts of the ontology, polarity and rating.

5. Conclusions and Future Works

The graphs produced for representing visually the collected data from online reviews from SMHs will help us to elicit better the visual requirements of our end-users. They will help us to know better some main aspects that may impact on the visualization design and to identify its formal requirements. In special, they will allow us to evaluate if a tree oriented visual representation of the data promotes an effective insight to the accommodation managers. However, there is much room for extending the visualization design of the data, improving its analysis and introducing new visual cues.

As a future works, we intend to perform user evaluations (visual response), to redesign these preliminary proposals according to the manager's needs and to proceed with the implementation of the visual layer of our tool.

Acknowledgements

References

[1] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu. OpinionSeer: Interactive Visualization of Hotel Customer Feedback. IEEE Transactions on Visualization and Computer Graphics, 6, 1109-1118. Nov-Dec 2010.

[2] C. Weaver, D. Fyfe, A. Robinson, D. W. Holdsworth, D. J. Peuquet and A. M. MacEachren. Visual Exploration and Analysis of Historic Hotel Visits. Information Visualization, 6:1, 89-103, February 2007.

[3] Rueijiau Sung, Chaochang Chiu, Peiyu Hsieh, Huiling Chou. In *Proceedings of the International Conference on Financial Management and Economics 2011*. Hong Kong. Vol. 11, 150-154. ISBN 978-981-08-9084-1.

[4] Michelle L. Gregory, Deborah Payne, David McColgin, Nick Cramer, Douglas Love. In *Proceedings of the International Conference on Weblogs and Social Media - ICWSM 2007*. Colorado, USA.

[5] Bross, J.; Schilf, P.; Meinel, C. Visualizing Blog Archives to Explore Content- and Context-Related Interdependencies. In *Proc. of the Web Intelligence and Intelligent Agent Technology (WI-IAT) IEEE/WIC/ACM International Conference on*. Toronto, Canada. 647-652, ISBN 978-1-4244-8482-9. September 2010.

[6] Heer, J; Bostock, M.; Ogievestsky, V. A Tour through the Visualization Zoo. Magazine Communications of the ACM, 53(6), New York, USA, 1-22, June 2010.

[7] L. Tweedie. Characterizing interactive externalizations. In *Proceedings ACM CHI*, p. 375–382, 1997.

[8] M. O.Ward and J. Yang. Interaction spaces in data and information visualization.In Proc. Eurographics/IEEE TCVG Symposium on Visualization, p. 137–145, 2004.

[9] J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. IEEE Transactions on Visualization and Computer Graphics, 13(6):1224–1231, 2007.

[10] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. IEEE Transactions on Visualization and Computer Graphics, 13(6):1216–1223, 2007.

[11] N. Elmqvist and P. Tsigas. A taxonomy of 3D occlusion management for visualization. IEEE Transactions on Visualization and Computer Graphics, 14(5):1095–1109, 2008.

[12] K. Brodlie. A classification scheme for scientific visualization. In R. A. Earnshaw and D.Watson, editors, Animation and Scientific Visualization, p. 125–140. Academic Press, 1993.

[13] S. Card and J. Mackinlay. The structure of the information visualization design space. In *Proc. IEEE Information Visualization*. Phoenix, USA, pages 92–99, 1997.

[14] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman. Do you see what I mean? IEEE Computer Graphics and Applications, 25(3):6–9, 2005.

[15] J. Bertin. Graphics and Graphic Information Processing. Walter de Gruyter, 1981.

[16] J. Bertin. The Semiology of Graphics. University of Wisconsin Press, 1983.

[17] W. S. Cleveland and R. McGill. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. J. of the American Statistical Association, 79(387):531-554, Sep. 1984.

[18] L. Wilkinson. The Grammar of Graphics. Springer-Verlag, 1999.

[19] Tableau.http://www.tableausoftware.com/products/desktop (visited on 25-12-2011).

[20] Tulip. http://tulip.labri.fr/TulipDrupal/ (visited on 25-12-2011).

[21] Treemap. http://www.cs.umd.edu/hcil/treemap/ (visited on 25-12-2011).

[22] Chaves, M. S.; Gomes, R. and Pedron, C. Analysing reviews in the Web 2.0: Small and Medium Hotels in Portugal. Tourism Management, Elsevier Science, 2011.

[23] Chaves, M. S. and Trojahn, C. Towards a Multilingual Ontology for Ontology-driven Content Mining in Social Web Sites. In *Proc. of the ISWC 2010 Workshops*. Vol. I, 1st International Workshop on Cross-Cultural and Cross-Lingual Aspects of the Semantic Web. Shanghai, China, Nov 7th, 2010.

[24] Godes, D. and Mayzlin, D. Using Online Conversations to Study Word of Mouth Communication. Market. Science, 23(4), 545-560, 2004.

[25] Turney, P. D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL.* p. 417–424, 2002.

[26] Chaves, M. S.; Trojahn, C. and Pedron, C. A Framework for Customer Knowledge Management based on Social Semantic Web: A Hotel Sector Approach. In: CRM and the Social and Semantic Web: Enabling Cliens Conexus. Colomo-Palacios, R.; Varajão, J. and Soto-Acosta, P. (Eds.). p. 141-157, Hershey, PA: IGI Global, 2012.

[27] C. Plaisant. The challenge of information visualization evaluation. In *Proc. of the Conference on Advanced visual interfaces*. AVI '04, 2004. ISBN:1-58113-867-9.

[28] Chen, C. and Czerwinski, M. In Journal International of Human-Computer Studies - Empirical evaluation of information visualization. Academic Press, MN, USA, 53(5), Nov. 2000.

[29] Lam, H., Bertini, E., Isenberg, P., Plaisant, C. and Carpendale, S. Seven Guiding Scenarios for Information Visualization Evaluation. TR 2011-992-04, Department of Computer Science, University of Calgary, 2011.