Capturing and Reusing Empirical Visualization Knowledge

Martin Voigt[1] and Martin Franke[2] and Klaus Meißner[1]

[1] Senior Chair of Multimedia Technology
[2] Software Engineering of Ubiquitous Systems,
Technische Universität Dresden
Dresden, Germany
martin.voigt,martin.franke,klaus.meissner@tu-dresden.de

Abstract. The context-aware discovery and ranking of visualization components is a crucial part of an adaptive, automated information visualization system. Since existing approaches allow for using expert knowledge formalized a priori, insights gained during the visualization processes by the users, e. g., suitable data-visualization combinations, are mostly neglected. In this paper, we propose a concept to capture and formalize these insights. Furthermore, we enhance a context-aware ranking approach using this knowledge by applying the well-known collaborative filtering. Thus, we are able to employ ratings also if data-visualization combinations are new for the current user.

Keywords: visualization, recommendation, adaptation, knowledge, collaborative filtering

1 Introduction

Due to the inexorable growth of data in all areas of life, humans face more and more the problem to understand the datasets they are confronted with. Thus, information visualization (InfoVis) tools are required to assist end users who are not familiar in creating effective graphical representations. Towards such an application tailored to visualization novices, we already proposed a visualization ontology (VISO) [8] to capture expert knowledge of the interdisciplinary domain, e. g., about data, graphic vocabulary, or human activity. This formal model is the foundation of a semantics-based InfoVis process [12] which guides the end user from identifying interesting parts of a dataset, over the context-aware selection of visualization components to the final configuration and perception of the visual representation. Its core is a discovery algorithm [11] to match the selected data but also to rank the suitable components according to factual visualization, domain, and context knowledge. Since the latter also includes users explicit preferences stored in a user model, we understand that it is hard for novices to explain which visualization techniques they like. Furthermore, we figured out that the knowledge created within the visualization process, i. e., which data-visualization combinations are appropriate, or inappropriate, in the current context, is lost so far.

Hence, our goal is not to neglect this kind of knowledge anymore and to use it for a better adaptation of the ranking of suitable visualization components. Therefore, we had to answer mainly two questions. Firstly, how could we extract the knowledge overburdening the end user? Secondly, how could we store, formalize, consolidate, and reuse the assembled information within our semantics-based visualization process? Thus, we present our solutions to tackle the mentioned problems by adapting concepts from the area of knowledge-assisted visualization as well as the domain of recommender systems. The concrete contributions are twofold: First, we propose a concept to *externalize* users visualization-specific insights by using implicit and explicit ratings, which are stored in a semantic knowledge-base. Second, we adapt our ranking algorithm to employ this empirical knowledge using collaborative filtering which allows for the indirect *collaboration* of end users within the visualization system.

In the following, we discuss the related work (Sect. 2), present our concept (Sect. 3), and finally conclude with some findings (Sect. 4).

2 Related Work

The work related to ours comes from the broad field of automated visualization systems. Here, the first sophisticated concepts and tools arose over 20 years ago. Unfortunately, *adaptation* of these systems, for instance to user needs and the device, has attracted less attention. With this regard, Golemati et al. [3] proposed a concept for a visualization environment for document collections taking into account user's, system's, and document's context. Similar to our work, they allow for an explicit rating of visualizations in the current context. This empirical knowledge is used for adaption in upcoming visualization selections of this user.

In contrast, Gotz et al. [4] and Nazemi et al. [5] present adaptive InfoVis systems building primarily on tracking users interactions and thus on their implicit ratings. Based on the interaction patterns they infer users task, match if the current presentation fits well or not, and if necessary adapt the graphic representation. These interaction-driven approaches leverage implicit state information, but they consider neither task information or preferences that are explicitly expressed by the user. Furthermore, in our concept the user's knowledge is reused in a collaborative manner as a global shared knowledge.

Since the adaption builds on many interrelated parameters, the knowledge-management within automated InfoVis tools is crucial. Thus, we rely on concepts from the domain of *knowledge-assisted visualization* as well. Its goal is to overcome the burden of learning complex visualization techniques by formalizing and sharing domain and visualization knowledge [1]. Wang et al. [13] propose a knowledge conversion process in visual analytics system. It comprises the *internalization* of knowledge of the user based on the graphical representation, the *externalization* of users insights based on interactions, or the direct *collaboration* of users during the visualization to foster the internalization. Bearing this theoretical framework in mind, we present an adaptive visualization system building on expert and empirical knowledge modeled using standardized ontology languages to foster their reuse and sharing.

As mentioned above, collaboration is a helpful concept to support users within the visualization process. Hence, intelligent InfoVis systems building on formalized expert and empirical knowledge may help. While the first could be captured and modeled a priori, the latter needs to be tracked, formalized, and reused on runtime. With this in mind, *collaborative filtering* is a well-known approach [9] which investigates similarity of ratings for items, in our case visualizations, given by users. Hence, no content analysis or tagging by experts is required. To the best of our knowledge, collaborative filtering has not been applied within automated visualization systems so far. In our opinion, the main reason is that they may identify suitable graphical representations based on users ratings but do not allow for an automated mapping of data to visual objects. However, if this matching is already provided by the visualization system, collaborative filtering introduces a new facet for rating and recommending visualization techniques: users conclusion if a mapping is suitable in the current context. Thus, it allows for tracking and storing user knowledge, enables the sharing of insights and the adaption of the visualization process due to this empirical knowledge.

3 Concept

Building on a context-aware recommendation algorithm from [11], which employs VISO [8] as ontological knowledge-base, in the following, we present concepts to externalize and reuse empirical knowledge evolving in the visualization process.

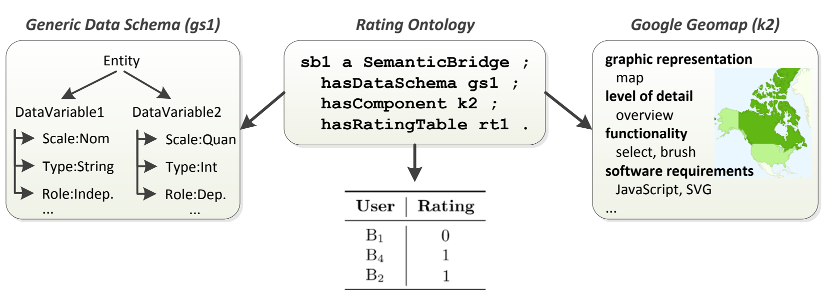

Fig. 1. Overview of the function of the rating ontology.

3.1 Externalization of Empirical Knowledge

The *externalization* process describes the storage of user's internal knowledge within the system. This process can be distinguished in acquiring *implicit* and *explicit* knowledge. The tracking and interpretation of interactions as knowledge is called *implicit rating* and is used for gathering knowledge, before the user marks down the end of the adaption loop by an *explicit rating*. In our concept, such a rating is always saved for a combination of a generic data schema and concrete visualization component using the *rating ontology* which semantically links both resources, see Fig. 1. The data schema is generated during the matching [11] and is an abstraction of the data a user has chosen for visualization including meta information like the scale of measurement or the data type for each data variable. The description of a visualization component holds all semantical information, e. g., its kind of graphic representation or interaction abilities. Finally, the ontology comprises the ratings of the users for a data-component combination. Based on this semantic interlinking, we are able to query for instance "only good rated components which visualize data on a map and allow for linking and brushing".

As mentioned above, we distinguish between explicit and implicit rating. To harvest implicit ratings, we rely on three actions from [6].

- **Repeated Use:** A visualization component is *used* more than three times by the same user. The usage of the component is recognized by doing a specific count of interactions or within a defined time interval. Since the repeated use is a sign that the user favors a component, it is temporary assigned with $r_{user} = 0.75$
- **Glimpse:** If the chosen visualization component is discarded without reaching a defined time interval or a count of interactions for recognizing the *Repeated Use*, it is downgraded and temporary assigned with $r_{user} = 0.25$
- **Related Rate:** The visualization component was explicit rated by the user, but in a different data combination. Since the user knows its characteristics, it is possible that he likes it for other data selections, which are distinct from the generic data schema, as well. In case of a good rating, the temporary is $r_{user} = 0.75$, otherwise $r_{user} = 0.25$.

Beside this implicit knowledge capturing, we gather *explicit* ratings. Thus, the user can explicitly decide whether the visualization is applicable for its purpose or not. Since we like to stimulate the user to rate, we employ a simple scale of *applicable* (1) or *not applicable* (0). We fulfill thereby the requirement to give the user an adequate possibility to rate, without an excessive demand.

3.2 Indirect Collaboration

The *collaboration* process describes a direct cooperation of two or more users [13], such as a chat or co-browsing. We broaden this scope by including indirect sharing of knowledge between users as collaboration. For our purpose, we need to reuse the collected knowledge within our ranking algorithm presented below. It comprises originally three kinds of ratings [11] according the factual visualization knowledge (r_{v_i}), the quality of the domain assignments (r_{d_j}), and the context knowledge (r_{c_k}). Now, it is extended by the user rating (r_{user}) to employ the user-generated visualization knowledge by using collaborative filtering. Thus, if the user has not rated the visualization component in the specific combination with a dataset, the algorithm tries to foresee a possible rating.

The combination of these factors is done using an arithmetic mean. The overall rating R which is calculated in terms of

$$R = \frac{1}{3|4}\left(\frac{1}{x}\sum_{i=1}^{x} r_{v_i} + \frac{1}{y}\sum_{j=1}^{y} r_{d_j} + \frac{1}{z}\sum_{k=1}^{z} r_{c_k}\ (+\ r_{user})\right)$$

has, therefore, a range between 0 and 1. We weight all three, respectively four, rating types equivalently for two reasons. First, the assignment of a (quantitative) rating is often subjective. Second, a complex user study is needed to evaluate the impact of each knowledge base in users visualization selection process what will be future work. The meaning factor $1/n$ is assigned with 1/3, if neither a user rating can be found, nor can be calculated. In all other cases it is assigned with 1/4 to keep the equivalently rating of all factors.

	$\mathbf{K}_1$	$\mathbf{K}_2$	$\mathbf{K}_3$	$\mathbf{K}_4$	$\mathbf{K}_5$
B_1	1	0	-	0	0
B_2	0	1	0	1	1
B_3	0	r_{rs}	1	1	1
B_4	1	1	1	1	-

Fig. 2. Example set for CFRS prediction

To identify an appropriate collaborative filtering algorithm, we analyze them by taking the accuracy, efficiency, stability, justification, and serendipity into account [2]. Hence, we decided to employ the item-based approach for our use case. It calculates the similarity between the ratings of visualization components, given by different users. With this information, the algorithm can predict the rating for the current visualization. An example is given in Fig. 2. It has to calculate the prediction r_{rs} for user B_3 and visualization component K_2. In this setting, the prediction is assessed based on the rating distances between K_2 and all other visualization components (K_1 - K_5). The algorithm chooses K_4 and K_5 as *nearest neighbours* and forecasts a rating of $r_{rs} = 1$. The formal calculation, which could be used without adaption, is considered in [10].

3.3 Architecture

To realize the concepts discussed above, we specified an architecture shown in Fig. 3 . It comprises three layers: ontological knowledge bases, loosely-coupled web services, and a component-based user interfaces. The first layer consists of three interconnected knowledge-bases. The *VISO* ①, contains all visualization specific knowledge, e. g., to describe the interaction and visualization capabilities of components. The API and meta information apart from the InfoVis domain of a component is described by the *Mashup Component Description Ontology* (MCDO) ② [7] which links to VISO concepts. As mentioned in Sect. 3.1, we designed an ontological knowledge base to store user's *ratings* ③ for the mapping of selected data to the chosen component. Therefore, it refers to concepts of the

VISO and MCDO. We build on different web service which heavily make use of the mentioned knowledge bases. The *Component Repository* ④ allows for the semantic-driven management of visualization components based on the MCDO. Further, the recommendation algorithm for appropriate components [11] is integrated within this service. Hence, it gathers amongst others the assessments stored or calculated within the *Rating Repository* ⑤ according to our proposed concepts. The visualization workflow of a user is accomplished by a composite web application called *VizBoard* ⑥ [12]. For example, it allows for searching for graphical representations, to integrate suitable ones and, finally, represented the data. At this last stage, it enables to explicitly and implicitly acquire users knowledge, like explained in Sect. 3.1, and save it using the Rating Repository.

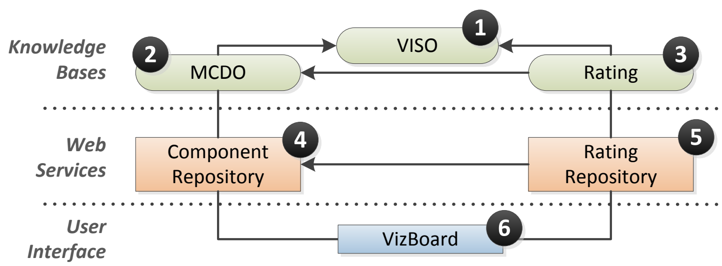

Fig. 3. Overview of the software architecture.

4 Conclusion and Further Work

The reuse of empirical knowledge evolving in a visualization process, especially proper mappings of data to visualization components within a context, was neglected and got lost so far. Hence, we proposed a concept for its externalization to capture, formalize and integrate the insights in an existing expert knowledge base. Furthermore, it becomes an essential part within our context-based ranking approach. Due to the application of collaborative filtering, we are able to employ explicit and implicit ratings also if data-visualization combinations are new for the current user. To the best of our knowledge, our approach is the first which employs formalized, inferred expert knowledge but also empirical, evolving knowledge from users to identify the most suitable visualization components.

Currently, we are also planning to conduct an exhaustive user study to identify and model the interdependencies between the knowledge bases employed within the ranking. Furthermore, we are working on a concept to use the a priori and empirical knowledge to assist the user in interpreting the visualized data what will underpin the usefulness of knowledge-assisted visualization.

Acknowledgements

This work is partly funded by the German Federal Ministry of Education and Research (01IA09001C); and by the European Social Fund and the Federal State of Saxony in Germany within project VICCI (ESF-100098171)

References

1. Chen, M., Hagen, H.: Guest editors' introduction: Knowledge-assisted visualization. Computer Graphics and Applications, IEEE 30(1), 15 –16 (jan-feb 2010)
2. Desrosiers, C., Karypis, G.: A comprehensive survey of neighborhood-based recommendation methods. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 107–144. Springer US (2011)
3. Golemati, M., Halatsis, C., Vassilakis, C., Katifori, A., Lepouras, G.: A context-based adaptive visualization environment. In: Information Visualization, 2006. IV 2006. Tenth International Conference on. pp. 62–67 (July 2006)
4. Gotz, D., Wen, Z.: Behavior-driven visualization recommendation. In: IUI '09: Proceedings of the 13th international conference on Intelligent user interfaces. pp. 315–324. ACM, New York, NY, USA (2009)
5. Nazemi, K., Stab, C., Kuijper, A.: A reference model for adaptive visualization systems. In: Proc. of the 14th Int. Conf. on Human-computer interaction: design and development approaches - Volume Part I. pp. 480–489. HCII'11, Springer (2011)
6. Nichols, D.M.: Implicit rating and filtering. In: In Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering. pp. 31–36 (1998)
7. Pietschmann, S., Radeck, C., Meißner, K.: Semantics-Based Discovery, Selection and Mediation for Presentation-Oriented Mashups. In: Proc. of the 5th Int. Workshop on Web APIs and Service Mashups. ACM (2011)
8. Polowinski, J., Voigt, M.: VISO: A shared, formal knowledge base as a foundation for semi-automatic infovis systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2013 (CHI'13) (2013)
9. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Recommender Systems Handbook, pp. 1–35. Springer (2011)
10. Sarwar, B., Karypis, G., Konstan, J., Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th international conference on World Wide Web. pp. 285–295. WWW '01, ACM, New York, NY, USA (2001)
11. Voigt, M., Pietschmann, S., Grammel, L., Meißner, K.: Context-aware recommendation of visualization components. In: Proc. of the 4th Int. Conf. on Information, Process, and Knowledge Management (eKNOW 2012) (2012)
12. Voigt, M., Pietschmann, S., Meißner, K.: A semantics-based, end-user-centered information visualization process for semantic web data. In: Hussein, T., Paulheim, H., Lukosch, S., Ziegler, J., Calvary, G. (eds.) Semantic Models for Adaptive Interactive Systems, pp. 81–106. Springer Berlin / Heidelberg (2013)
13. Wang, X., Jeong, D.H., Dou, W., Lee, S.W., Ribarsky, W., Chang, R.: Defining and applying knowledge conversion processes to a visual analytics system. Computers & Graphics 33(5), 616 – 623 (2009)