Semantic Support for Visual Data Analyses in Electronic Commerce Settings

Jens Gulden[(✉)]

University of Duisburg-Essen, Universitätsstr. 9, 45141 Essen, Germany
`jens.gulden@uni-due.de`

Abstract. While the value of visualizations for understanding and exploring knowledge is considered high in diverse fields of application, the efforts for creating effective and efficient data visualizations often outweigh the capacities of individuals and organizations to create their own data visualizations from scratch. Hence, software tool support is demanded to allow users who are not experts in creating visualizations to have access to these visual means of expression as well.

The question whether a data visualization is "good" in the sense of whether it can fulfill the information needs of involved stakeholders, however, highly relies on an understanding of the way domain stakeholders view the available information and ask questions about it. This semantic aspect of data visualization is not explicated by existing approaches for data visualization development. The following article proposes a methodical approach which explicates knowledge about the meaning of data in the form of conceptual models, and interweaves the creation process of visualizations with an analysis of the information needs of involved domain stakeholders. An exemplary application of the method in the e-commerce domain is included.

1 Methodical Support for the Creation of Data Visualizations

The use of data visualizations in information systems increasingly attracts interest in science and practice [6,11,15]. Especially in the area of electronic commerce (EC), which inherently combines business scenarios with underlying automation facilities [17,19,20], rich sets of data are generated during every-day operation. They cover a broad range of semantics, are available at large volumes, and reach a comparably high data quality due to the high degree of automation in their creation. In order to retrieve knowledge from such amounts of data, visual analyses tools are regarded as efficient means for humans to aggregate, combine, and navigate data interactively [8,18].

The human mind generally performs multiple cognitive actions in parallel, on diverse levels of detail and granularity. Presenting data visually, and preparing the visualization in a way that relevant relationships and facts become perceivable as simultaneous elements of a rich information environment [9], is one next consequent step in increasing the quality of existing EC systems. It is thus

desirable to support involved groups of user, e.g., analysts, managers, and data scientists, with methodical support to perform these analyses efficiently in terms of invested efforts for creating visual representations, and effectively with respect to the goal of understanding and / or discovering business relevant knowledge in the data.

The examination presented in this paper proposes a method for suggesting data visualizations based on the domain-specific information needs of stakeholders specifically in EC analysis settings. To achieve this, semantics of the specific domain of electronic commerce is incorporated with the help of enterprise models [12,16].

2   Related Work

A few approaches have been proposed that offer basic support to generate visual representations from business data [5,10], and diverse software products offer visualization wizards which allow for automatic visualization generation [1–4].

These approaches make use of syntactic features of the available data, and typically propose a set of possible visual representations and navigation options that can validly be constructed from the syntactic features. The drawback of such approaches is that depending on the complexity of the underlying data, there may be many syntactically possible forms of visualizations and navigation options, which from a semantic point of view make no sense or are even counter-productive to be offered in an analysis environment. This is because the visualization mechanisms are agnostic towards domain-specific semantics associated with the available data.

It would instead be desirable that the creation of visualizations for data analysis could exploit knowledge about the semantic domain from which the data originates, so it would be better able to propose which combinations of data to visualize, and which visual expression means are best suitable to fulfill information needs relevant to the given domain.

A central deficiency of visualization methods that follow state-of-the-art approaches, is that they typically offer direct mappings from available data to visual means of expression only [15]. The major software products for data visualization [1–4] seem to compete in providing growingly complex mapping types and extensive libraries of presets and templates, but the fundamental problem remains that for each visualization type to be created, the mappings have to be decided ad-hoc over and over again. Means for expressing design decisions about *why* mappings have been chosen, are not systematically integrated into the applied visualization method. As a consequence, it is typically not possible to reuse design rationales, since they are not recorded nor reproducable in a systematic way.

3   A Method for Visual Data Analyses with Semantic Support in Electronic Commerce Settings

Using information systems in electronic commerce is a special case of using information systems in general, with the EC domain determining a special contextual

realm. This domain comprises of specific stakeholders (e.g. *seller*, *buyer*), objects (e.g. *product*, *catalog*, *sale*, *invoice*), and processes (e.g. *order*, *delivery*, *return*). Especially with regard to the processes and structures behind EC interactions, specific information needs of the involved stakeholders can be identified for this domain, which make it possible to perform an in-depth analysis of requirements towards the analysis demands in EC that wouldn't be possible to perform on a general domain-independent level of reflection.

It is central to the suggested approach, that instead of mapping available data directly onto input variables for visualizations, it gets associated with conceptual models that describe selected EC scenarios and corresponding information needs. The conceptual models in turn are associated with visualizations that are suitable for fulfilling these information needs. Conceptual models that describe business-scenarios in such a domain-specific way are called enterprise models [12]. The involvement of enterprise models shifts the expressiveness of the suggested approach onto a level of economic meaning of data, rather than operating on a purely syntactic level of matching combinations of available values to input types of visualizations.

For the end-user, the application of the approach follows this procedure: Available data sources of EC systems are described using a standard data model, as it can be exported, e.g., by a relational database management system. With this model as input, a configuration wizard can iterate over all elements of the previously developed conceptual enterprise models and ask the user, which elements of the input data should serve as instances of the domain-specific types described in the conceptual models. Interactively, the user associates the available data with elements of the conceptual models, thus defines their domain-specific semantics.

3.1 Requirements Towards an Approach for Domain-Specific Data Visualization

This section shortly discusses 5 main requirements towards an improved methodical approach for visualization specifications in e-commerce.

Req. 1: Explicate the meaning of data as part of the method. As argued, assuming a purely syntactic relationship between data and possible visualizations thereof does not offer a distinctive enough basis for making automatic suggestions for visualizations types. The larger the number of input variables, and the more complex the relationship structures among data sources get, the less effective automatic suggestions can be made which fulfill information needs of domain stakeholders. As a consequence, a higher degree of expertise and experience is required for choosing meaningful visualizations among automatically provided suggestions, which in turn limits the amount of users who can create visualizations on their own, and the range of applications where visualizations can be used in an economically reasonable way.

A method for creating data visualizations should thus incorporate means for explicating domain-specific semantics, i.e., describe the meaning of data with the help of conceptual models, as a basis for more focused automatic means for suggesting meaningful visualizations.

Req. 2: Support identification of information needs. To identify information needs of involved stakeholders, an understanding of the domain from which data to be visualized originates is required, both independent from any actually available data, as well as from possible visualization options.

The use of conceptual models, e.g., enterprise models for business related domains, allows to point out relevant objects of interest for visual analyses as elements of conceptual models, which themselves can have a visual notation which connects to visual metaphors known in the modeled domain. This way, as an initial part of the method, domain experts can consciously negotiate on which information elements to put in focus of an analysis.

Req. 3: Justify meaningful visual means of expression. It is a consequent next step after having identified the stakeholders' information needs, to choose visual means of expression that can fulfill these needs effectively and in a cognitive efficient way. This requires a high level of design expertise, and is thus a task that should be performed by specially trained experts as part of the preparation phase of the method. It is possible to shift this into the preparation phase, because specific domain knowledge is already available at this point, and the method allows to explicate the meaning of data constructs (see Req. 1).

Req. 4: Enable reuse of domain-specific visualization types. The responsibility for justifying choices of visualization types should lie in the initial preparation phase of the method, where domain experts and visualization designers consciously reflect on the use of visualization types for the purpose of fulfilling information needs.

A visualization method should thus support the specification of visualization types in relation to semantic domain concepts in the preparation phase, and later allow to reuse these specifications by an end-user in concrete application contexts.

Req. 5: Provide automatic guidance in creating visualizations. Finally, the target requirement is the demand for efficient and effective automatic guidance in creating visualization that answer stakeholders' information needs and allow to explore existing data from the relevant perspectives of domain experts. The fulfillment of this requirement is achieved when all previous ones are fulfilled; in this sense, this last requirement subsumes the previous ones and represents the overall goal to develop an effective and efficient visualization method based on semantic characteristics of a domain.

3.2 Method Architecture

The method elaborated in the following suggests the use of enterprise models as a semantic intermediate layer when defining connections between data and meaningful visualizations. This way, enterprise models take in the role of a semantic repository, which allows to systematically express how meaning of data is reflected through visual representation means in the range of a given domain. The building blocks of this methodological architecture, in contrast to direct mapping approaches, are depicted in Fig. 1.

3.3 Procedural Steps and Involved Roles

The method consists of 5 steps, which are shown in List 1 for an overview, and are described in depth in the following sub-sections. The procedure is divided into an
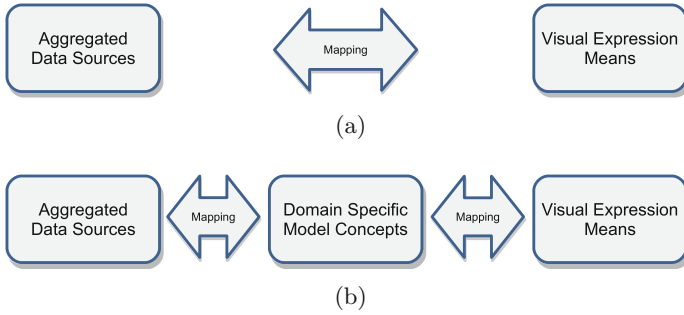
Fig. 1. Overview on the (a) traditional direct mapping architecture for visualizations, compared to the (b) approach suggested in this paper

initial sequence of steps which are performed once to configure the method for a specific domain. This is done by expert method engineers, together with domain experts from the domain in focus and visualization designers. When applying the configured method, a role change can take place, which allows any involved domain stakeholder to define concrete visualizations for particular instances of the domain, based on re-usable definition previously defined in the configuration steps. The different responsibilities throughout the process and a possible point for division of labor is symbolized by a horizontal line between steps 3 and 4.

Step 1: Explicate relevant semantic concepts of the examined domain
Step 2: Derive domain-specific information needs from the modeled scenarios
Step 3: Design visual expression means to fulfill the identified information needs in an effective and efficient way

———————

Step 4: Associate available data from operative systems based on the semantic models
Step 5: Select visualization types and navigate data to achieve concrete data visualizations

List 1: Steps of the suggested method

This general conceptualization of a semantics-based visualization method is applied to the e-commerce domain in the following section.

4   Application to the E-Commerce Domain

To further describe a concrete application scenario of the method, this section puts its focus on exemplifying the adaptation phase of the method to the e-commerce domain, which is performed in the methodical steps 1 to 3 (Sect. 3.3). This is done by first introducing fundamental concepts of the e-commerce domain

with models of a prototypical e-commerce setting (step 1), then deriving domain-specific information needs from the modeled setting (step 2), and finally designing justified effective and efficient visualizations (step 3). The end-user application steps 4 and 5 complete the method description.

4.1 Step 1: Model Typical Business Scenarios of the Examined Domain

The method preparation starts by capturing domain-specific processes and structures with multiple interconnected enterprise models [12,16]. As an example, the core business process model of a general e-commerce shop is shown in Fig. 2. The business process starts with a customer placing an order via the internet, which is shown on the left-hand-side of the business process model. Subsequent tasks for processing the order are shown in left-to-right order, with black lines between them indicating the control flow of the process. The model also contains references to involved actors, as well as to resources of diverse kinds [13].
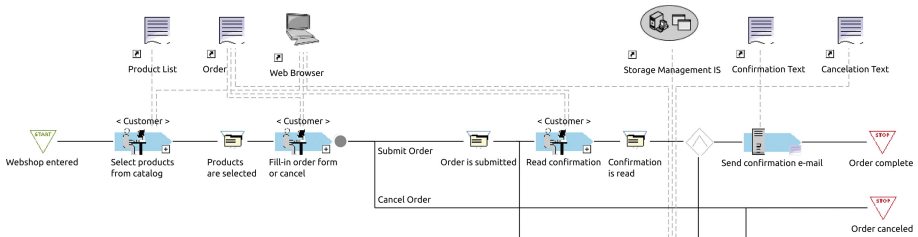
Fig. 2. Excerpt of the business process model of an online order process

Figure 3 shows an exemplary model of the organizational structure of actors and a model of resources.

These domain-specific models carry a high degree of semantics in their conceptual elements. Unlike with models in general purpose modeling languages, which intentionally provide highly general concepts such as *Object* or *Relationship*, each single model element in domain-specific models can be interpreted deeply on the basis of domain knowledge about it. This is, e.g., the case with the modeled concept of a *customer* actor, and resources such as *products* and *product-lists*. With the domain-specific knowledge attached to these concepts, and their contextual settings explicated in the conceptual models, specific information needs and justified domain-specific analytical questions towards the available data can be formulated in the following step.

4.2 Step 2: Derive Domain-Specific Information Needs from the Modeled Scenarios

Because the domain of the analysis setting is described on an abstract level by domain-specific models, classes of analysis questions can now be identified
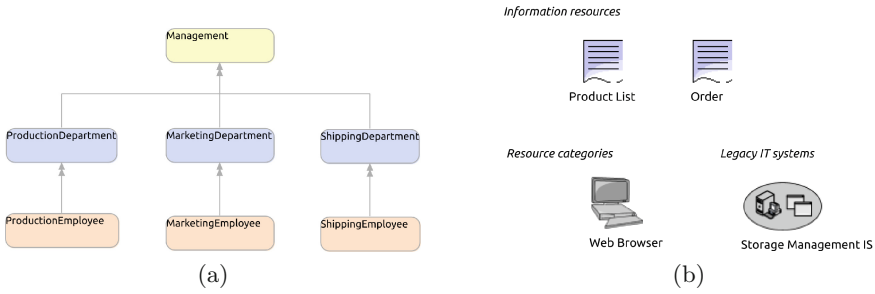
Fig. 3. Organization model (a) and resources model (b) according to the example process

which characterize information needs of the involved stakeholders. In e-commerce settings, it can generally be spoken of *products* that are offered through an automatic *catalog* mechanism to *customers*, with whom *sales* are performed, and sometimes *returns* of products occur. This conceptualization captures the domain of an e-commerce business on a general level, which makes it applicable to almost any concrete instance of actual e-commerce enterprises. However, despite its wide generality, the conceptualization still provides a rich body of semantics about the given domain for which meaningful visualizations are to be developed.

This degree of semantics can now be harnessed to anticipate general analytical questions towards the development and status of an e-commerce business. It becomes possible to explicate the information needs of involved stakeholders in the domain on the general level of questions that are induced by the semantic specifics of the domain.

In case of the e-commerce domain, the semantically rich basic concepts *products*, *catalog*, *customers*, *sales*, and *returns* can be brought in relation prior to creating any visualization by formulating domain-related analytical questions. These can operate with the specific semantics of the basic concepts, e.g., the assumptions that relationships between customers and products they buy are of specific interest for some stakeholders, and that both products and customers can be categorized to customer-groups and products-groups. This allows to formulate detailed, yet re-usable, analytical questions towards the status and development of an e-commerce business. The validity of these questions can be evaluated through professional discourse among domain experts without the need for expertise in information visualization.

An initial set of questions for the e-commerce domain is suggested in List 2, which take in the perspective of owning and operating stakeholders of e-commerce businesses. Naturally, this list is not finite and can be extended whenever additional information needs are identified by domain experts.

– Are there any products, which are particularly attractive / unattractive for specific customers or customer groups?
– How have sales or returns of particular products and / or product groups developed over a given period of time?

– How frequent and at which volumes do selected customers and / or customer groups place orders?
– Are there products and product groups which cause unexpected amounts of product returns?
– Are there periods of time in which particular customers or customer groups show different order / return behavior than usual?
– How is the relationship between views of products in the product catalog compared to actual sales of these products?
– ...

List 2: Analytical questions derived from the domain models

The analytical questions derived on an abstract level can now be examined by experts for visual information representation, and default visualization types can be developed specific for the information needs in the examined domain. This is done in the following step for a sub-set of the above listed analytical questions.

4.3   Step 3: Design Visual Expression Means to Fulfill the Identified Information Needs

Based on the previously elaborated analytical questions, classes of meaningful visualizations are now suggested which serve the purpose to give cognitive efficient and effective insight into answers on analytical questions. It cannot be the aim of this paper to summarize the body of knowledge in the entire discipline of information and data visualization. Therefore, the actual process of deciding in detail which visual means of expression are suited best to serve the identified analytical purposes, relies on the expertise of visualization designers and literature from the data and information visualization domain, e.g. [6,7,11,18]. Examples of principles and best practices applied during this step are, e.g., that relations among values can be expressed well by projecting them onto a 2D plane (e.g. using scatter-plots of values), or that quantitative values are best represented by one-dimensional visual constructs. Based on these principles, for each previously identified analytical question (Sect. 4.2) one or more visualization types are now to be designed.

Visualization types for two of the identified analytical questions are sketched in Fig. 4 to demonstrate the applicability of the method. Figure 4 (a) depicts a scatter-plot diagram, which is suitable to provide insight into relationships between customers, products and sales figures, either for individual instances of the customer and product concepts, or for categorized groups formed out of individuals. This visualization type thus is suitable to fulfill information needs imposed by the analytical question "Are there any products, which are particularly attractive / unattractive for specific customer groups?" (Sect. 4.2). It provides a rich set of navigation options by selecting the entities to display, which makes it a powerful visual tool both for data explanation, as well as for exploration. Terms written in angle brackets "<" and ">" indicate placeholders for
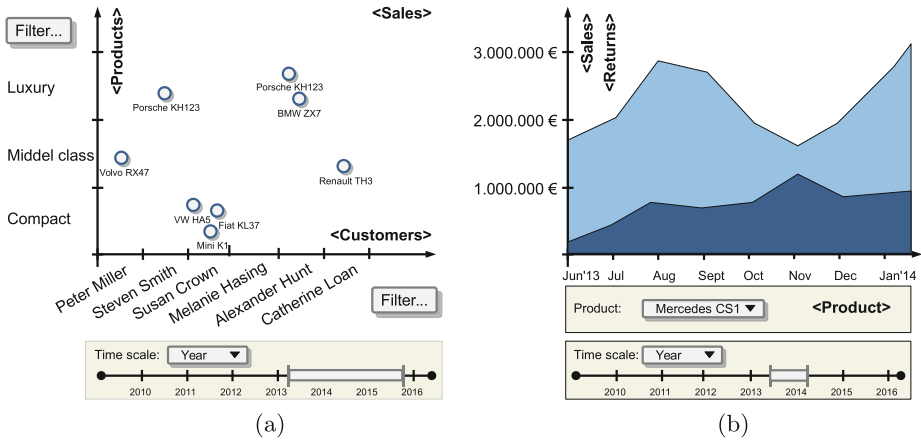
Fig. 4. Two example visualization types derived from analysed information needs

concepts from the domain model, which will need to be associated to concrete data sources, before a concrete visualization will be available.

Figure 4 (b) shows the sketch of a timeline diagram, which displays the development of sales versus returns in a configurable range of time, filtered by products, or their respective categories. This type of visualization is suitable to answer analytical questions such as "How have sales or returns of particular products and / or product groups developed over a given period of time?" (Sect. 4.2).

The actual values to render in instances of these diagram types are not known at this stage of visualization design yet. It will be up to the end-user to determine appropriate data sources which match the semantics of the modeled domain concepts (see the following Sect. 4.5). To give an impression of the utility of the visualization types, the sketches in Fig. 4 are displayed with example values that represent possible appearances of the visualizations.

4.4   Step 4: Associate Available Data from Operative Systems Based on the Semantic Models

At this point in applying the method, the user role potentially changes to be fulfilled by any domain expert and involved stakeholder with domain-specific information needs. This means, no specific competencies in designing visualizations, especially in justifying effectiveness and efficiency of data visualizations, is required for this and the next task. Expert knowledge can be reused that has been explicated in the earlier adaptation steps of the method.

In order to define actual data sources of e-commerce systems as sources for domain-specific data visualization, the user now associates data elements to concepts specified in the domain models. This can, e.g., be carried out by providing a mapping between domain model concepts to results of queries or views on a relational database. In such a setting, each row in a query result table

represents an instance of the associated domain model concept, table columns can subsequently be mapped onto attributes of the respective concepts.

Figure 5 exemplifies the integration of SQL query definitions to map between data sources and conceptual models elements into the graphical user interface (GUI) of the MEMOCenterNG enterprise modeling tool [14]. The process of associating database queries can be supported by a semi-automatic wizard, which iterates through all available conceptual model elements and guides the user to fill in the appropriate settings.
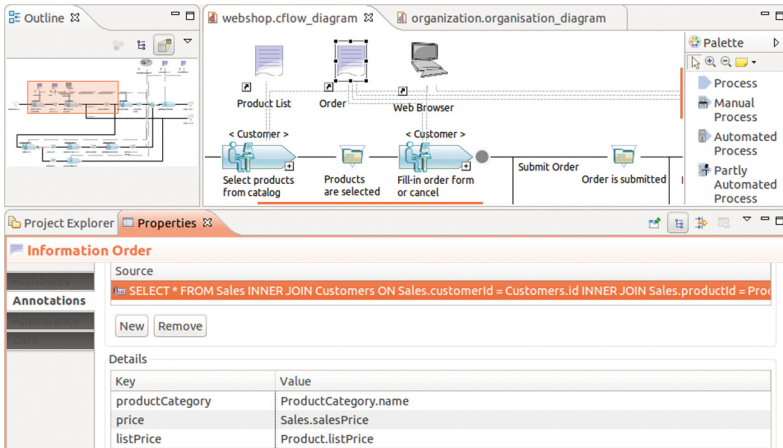

Fig. 5. Integration of SQL query definitions into MEMOCenterNG

As a difference to existing approaches, it is important to note that at this point the user does not have to think in terms of axes, intercepts, coordinates, colors, or any visual property of the visualizations to be created, as it is the case with existing approaches. Instead, data is mapped to concepts of the analysis domain, which have previously been adapted to justified visualization types. Via this intermediate semantic layer, the method increases effectiveness and efficiency of the proposed visualizations and becomes reusable for multiple cases of concrete applications.

4.5 Step 5: Select Visualization Types and Navigate Data to Achieve Concrete Data Visualizations

With a growing number of specified associations between data and conceptual models, gradually more analysis scenarios become accessible. A selection wizard can dynamically display the list of visualization types that are possible to be rendered on the basis of the currently available associations between data sources and model concepts. Interactivity features, such as navigation options and aggregation possibilities, can be derived from the semantic information about

temporal, spatial, and categorial relationships specified in the conceptual enterprise models.

The result are data visualizations for concrete analysis scenarios specified by the end-user, derived from visualization types with underlying justified design decisions, that ensure effective and efficient analyses in the specified domain.

5   Conclusion

With the results achieved, a software-implementable method is described which provides advanced tooling support for interactive visual data analyses for electronic commerce (EC), based on business semantics described with enterprise models. This allows to efficiently perform visual analyses of data in EC settings, and effectively gain business-relevant knowledge for this particular domain.

The requirements stated in Sect. 3.1 can be regarded as fulfilled by the proposed approach. *Req. 1: Explicate the meaning of data as part of the method* is fulfilled by using conceptual models as underlying explication of relevant domain concepts (Sect. 4.1). As argued, *Req. 2: Support identification of information needs* is subsequently fulfilled by incorporating a reflective analysis of stakeholders' views on the domain into the sequence of domain adaptation steps (Sect. 4.2). Shifting the responsibility for developing justified visualization types into the method's domain adaptation phase (Sect. 4.3), serves to achieve the purposes of *Req. 3: Justify meaningful visual means of expression*. The proposed interfacing description mechanism based on semantic concepts rather than only syntactic data characteristics (Sect. 4.4) makes the result of the adaptation phase applicable for later concretizations in practical settings as demanded by *Req. 4: Enable reuse of domain-specific visualization types*. By fulfilling these four key requirements, the subsuming *Req. 5: Provide automatic guidance in creating visualizations* gets fulfilled in total as well.

Future work will consist of elaborating a richer set of analytical visualization types and performing user studies based on prototypical tooling.

References

1. IBM Watson Analytics. http://www.ibm.com/analytics/watson-analytics/
2. Lyra Visualization Design Environment (VDE). http://idl.cs.washington.edu/projects/lyra/
3. SAP Lumira. http://saplumira.com/
4. Tableau Visual Analytics Software. http://www.tableau.com/
5. Al-Kassab, J., Ouertani, Z.M., Schiuma, G., Neely, A.: Information visualization to support management decisions. Int. J. Inf. Technol. Decis. Making **2**, 407–428 (2014)
6. Borner, K., Polley, D.E.: Visual Insights: A Practical Guide to Making Sense of Data. MIT Press, Cambridge (2014)
7. Cairo, A.: The Functional Art. Voices That Matter. Pearson Education, New York (2010)
8. Chen, C.: Information visualization. Springer, London (2010)

9. Cooper, A.: About Face: The Essentials of Interaction Design, 4th edn. Wiley, Hoboken (2014)
10. de Sousa, T.A.F., Barbosa, S.D.J.: Recommender system to support chart constructions with statistical data. In: Kurosu, M. (ed.) HCI 2014, Part I. LNCS, vol. 8510, pp. 631–642. Springer, Heidelberg (2014)
11. Few, S.: Information Dashboard Design: The Effective Visual Communication of Data. O'Reilly, Sebastopol (2006)
12. Frank, U.: Multi-perspective enterprise modeling (MEMO) - conceptual framework and modeling languages. In: Proceedings of the Hawaii International Conference on System Sciences (HICSS-35), Honolulu (2002)
13. Frank, U.: Multi-perspective enterprise modelling: background and terminological foundation. Technical report 46, ICB Institute for Computer Science and Business Information Systems, University of Duisburg-Essen, Essen, December 2011
14. Gulden, J., Frank, U.: MEMOCenterNG - a full-featured modeling environment for organization-modeling and model-driven software development. In: Soffer, P., Proper, E. (eds.) Proceedings of the CAiSE Forum Hammamet. CEUR Workshop Proceedings, Tunisia, 9–11 June, vol. 592, pp. 76–83. CEUR (2010). ISSN 1613–0073
15. Gulden, J., Reijers, H.A.: Toward advanced visualization techniques for conceptual modeling. In: Grabis, J., Sandkuhl, K. (eds.) Proceedings of the CAiSE Forum Stockholm, CEUR Workshop Proceedings, Sweden, 8–12 June. CEUR (2015)
16. Sandkuhl, K., Stirna, J., Persson, A., Wißotzki, M.: Enterprise Modeling - Tackling Business Challenges with the 4EM Method. Springer, Heidelberg (2014)
17. Shaw, M., Blanning, R., Strader, T., Whinston, A.: Handbook on Electronic Commerce. Springer, Heidelberg (2000)
18. Whitney, H.: Data Insights: New Ways to Visualize and Make Sense of Data. Morgan Kaufman, Waltham (2013)
19. Yokoo, M., Ito, T., Zhang, M.: Electronic Commerce: Theory and Practice. Springer, Heidelberg (2008)
20. Zhang, F.: The application of visualization technology on e-commerce data mining. In: Second International Symposium on Intelligent Information Technology Application, IITA 2008, vol. 2, pp. 563–566. IEEE (2008)