IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITP5177
This paper appears in *Managing Modern Organizations Through Information Technology*, Proceedings of the 2005 Information Resources Management Association International Conference, edited by Mehdi Khosrow-Pour. Copyright 2005, Idea Group Inc.

Exploration of Conceptual Models: Application of the GOM Framework

Christian Janiesch and Christian Brelage
European Research Center for Information Systems (ERCIS), Leonardo-Campus 3, 48149 Muenster, Germany,
{christian.janiesch, christian.brelage@ercis.de}

Roland Holten
Frankfurt University, Mertonstr. 17, 60054 Frankfurt, Germany, holten@wiwi.uni-frankfurt.de

ABSTRACT
Conceptual modeling is widely perceived as a core element of the IS discipline. Conceptual models are used for a variety of different purposes (for instance software engineering or organizational design). However, the quality and adequacy of conceptual models is difficult to access. The Guidelines of Modeling (GoM) are proposed as generalized criteria for measuring of model quality. However, the GoM framework has still to be instantiated in order to provide operational criteria for a specific modeling purpose. We contribute to this work by a formal exploration of conceptual models of web-based information systems. The analysis was found to be very useful for assessing a complex web-based information system with special regard to navigational issues ex post.

CONCEPTUAL MODELING IN THE IS FIELD
Conceptual models are created by using conceptual modeling grammars, for instance the Entity-Relationship Model grammar. Conceptual modeling grammars provide sets of constructs that can be used to depict phenomena within the domain of discourse. Since requirements vary greatly and are diffuse, conceptual modeling grammars are usually less strictly specified than formal grammars. Models created with these semi-formal languages provide clarification when depicting aspects from the domain of discourse. Being less formal, the same aspect may be depicted in several ways. Consequently, the quality of the model, in terms of the modeler's intention and the modeling objectives, is difficult to assess.

In an attempt to address this assessment issue, several concepts and guidelines have been proposed in order to improve the quality and adequacy of conceptual information models, for example the Guidelines of Modeling (GoM) proposed by Becker et al. [Bec+95]. They are intended as a generic approach to the improvement of conceptual models (refer to Section 2.1 for details on related work). Due to ambiguity and the broad variety of modeling tasks however, it is difficult to provide operational criteria (i. e., an instantiation of GoM) for modelers.

We contribute to this issue in the following way: We describe a case in which conceptual models were explored in order to gain insights in a web-based information system. The measures that were used can serve as functional criteria for conceptual models of this particular type of information system.

Our specific research questions (RQ) are:

RQ1: What insight can be gained from an analysis of conceptual models with respect to the adequacy of the current design? We provide empirical evidence from a case study conducted with a web-based information system. The implementation of the information system was modeled conceptually ex post. The models were explored in detail in order to derive bases for further improvement.

RQ2: Does the GoM-Framework provide the expected beneficial input? We use the GoM framework to evaluate the construction and application of the navigational structure model and, thus, are able to assess the adequacy of the model and thereby indirectly the quality of the original design of the information system's navigational structure.

RELATED WORK & RESEARCH APPROACH

Related Work
The quality of information models is of paramount importance for the efficient development and overall quality of any information system. There are various different frameworks for quality assurance and evaluation of information models, each with different focal points. Whereas the approaches of Batini et al. [Bat+92] and Moody & Shanks [MoSh94], [MoSh03] focus on data modeling, other approaches (e. g. Lindland et al. [Lin+94]) make use of high-level-statements that hinder a useful application in modeling projects or merely collect data from case studies [Maie96] without any attempt at generalization. Becker et al. introduced the *Guidelines of Modeling (GoM)* that postulate six principles which improve of the quality of conceptual models without the need to adhere to a specific modeling technique or framework like the Event-Driven Process Chain (EPC) or Entity-Relationship Model (ERM) [Bec+95]. Current proposals to evaluate the quality of conceptual information models assess the modeling technique or even the research approach, rather than the model itself (e. g. [Hev+04, Sha+03, Wan+99]).

Based on our objective to explore and evaluate mainly navigational structures and not the underlying modeling technique, a general approach to model evaluation is applied. Even though the GoM approach is not specialized in analyzing navigational structures, it was chosen because conceptual models are the domain of discourse and its adaptable form allows an application in this particular case.

Guidelines of Modeling
The aim of the GoM is the development of specific design recommendations, whose application increases the quality of models beyond the mere fulfillment of syntactic rules [Bec+95]. While the observance of the principles of correctness, relevance and economic efficiency is a necessary precondition for designers of models, the principles of clarity, comparability and systematic design merely have a supplementary character. Furthermore, different views and methods can be taken into account which facilitate the application to specific contexts such as the exploration of navigational structures.

The *Guideline of Correctness* has two dimensions: syntactic and semantic correctness. A model is syntactically correct if it is consistent and complete in terms of the meta model on which it is based. Semantic correctness requires that the structure and behavior of the model replicate the real world as intended by the modeler. In our case, semantic correctness refers to the consistent usage of names as well as an adequate structuring of information.

While other frameworks use "completeness" as a quality factor of information models [Bat+92, MoSh03], the GoM express this criteria in more relative terms. The *Guideline of Relevance* postulates that a relevant section of the real world is selected and that the chosen level of abstraction does not interfere with the model's understandability.

The *Guideline of Economic Efficiency* constrains all other guidelines. It is comparable to the "feasibility" criteria of Lindland et al. [Lin+94] and usually restricts the correctness and/ or the clarity of a model.

The pragmatic aspect of the semiotic theory is integrated in the GoM through the *Guideline of Clarity*. Without a readable, understandable, and comprehensible model, all other efforts are futile. However, this guideline is extremely subjective and postulates that the model is understood by the model user. The arrangement of the navigation elements on a website however, may still obliterate any clear structure if it does not follow common rules.

The *Guideline of Comparability* enforces the identical use of all guidelines. This guideline includes, e. g., the conform application of layout or naming conventions. Otherwise, two models would follow certain, but different conventions. The need to compare information models is obvious since a navigational structure must be comparable to those of other websites.

The *Guideline of Systematic Design* postulates well-defined relationships between information models that belong to different views. Therefore, one requirement is to use a meta model that integrates all the relevant views. In our case, only models of navigational structures are explored. Thus, the Guideline of Systematic Design does not apply.

EXPLORATION OF CONCEPTUAL MODELS

Description of the *asinfo* Case

The objectives of the *asinfo* project (http://www.asinfo.de/) were the identification and analysis of factors that anticipate the application and implementation of industrial health standards in organizations. The analysis revealed that organizations face several obstacles when attempting to satisfy their specific industrial health information needs. In particular, specific information is hard to find (since it is distributed among several providers), taxonomies vary (reflected in different key terms and navigational structures), and language problems between experts and non-experts hinder efficient communication.

The web-based information system *asinfo* provides a comprehensive overview of documents relating to the maintenance of industrial health, safety standards, and provisions on labor. The system integrates information from more than 130 websites of external information providers. Their documents are scanned on a regular basis by the system, indexed, and integrated in the system. The indexing algorithm works in the same way as common web search engines. *Asinfo* does not provide any information by itself. The system can be accessed without registration or any payment by the public and is, to the best of our knowledge, the most comprehensive library of its kind in Germany.

The objective of our work was to gain insight into the structure of *asinfo* by an intensive ex post analysis of *asinfo*'s navigational capabilities and the identification of problems as well as suggestions for further improvement of the system. Moreover, a detailed conceptualization of the system was developed in order to improve our understanding of *asinfo*'s navigational structure and its usage, so as to satisfy the information needs of users.

Conceptualization of *asinfo*

The conceptualization of *asinfo* is based on the work of Becker et al. [Bec+03a, Bec+03c] who developed a modeling technique for web-based information systems, which itself is based on work by Holten (e. g. [Bec+03b, Holt03]) who introduced a conceptualization for specifying management views in information warehouse projects. The extensive modeling technique provides concepts for describing and analyzing web-based information systems holistically and in detail.

In order to improve understanding of the conceptualization, some core concepts and their counterparts in *asinfo* are described briefly. *Content* represents information objects stored in the system. Content may be structured (*Content Structure*), reflecting its complex nature. The conceptualization corresponds to the hypertext paradigm (content as nodes, content structure as links between nodes).

Hierarchies have traditionally been used for structuring and classifying phenomena from the real or imaginary world. Due to their (technical) advantages and inherent concepts of aggregation and abstraction, hierarchies were investigated intensively in research on hypertext [FaFa00, LaCz98]. Moreover, hierarchies reflect an appropriate ordering of things in the world and are, if properly engineered, intuitively understandable for users. In order to adapt and extend the concept of hierarchies, the concept of *Reference Object* is introduced. This represents an object or concept from the real or imaginary world that is relevant in the given context of the information system. *Dimension Reference Objects* are a subset of reference objects are which are ordered hierarchically and characterized by an explicit assignment to exactly one *Dimension*. Furthermore, dimension reference objects can be divided into leafs and nodes. All content is assigned to at least one reference object, implying that this content is relevant whenever an individual wishes to retrieve information about the reference object. Thus, reference objects classify the documents. Dimensions are navigation trees (hierarchies) in web-based information systems. However, it is very useful to use more than one hierarchy, and as such using different classification schemes, in order to structure the information in the system. Thus, several dimensions may be used in a system, allowing alternative search paths for users. Moreover, information can be characterized or classified much more appropriately with more than one hierarchy. If alternative hierarchies are offered, users, especially non-experts who are not familiar with the expert terminology, may find information more efficiently. Taken together, these dimensions form a multidimensional space which we refer to as the *Information Space*, see Figure 1.

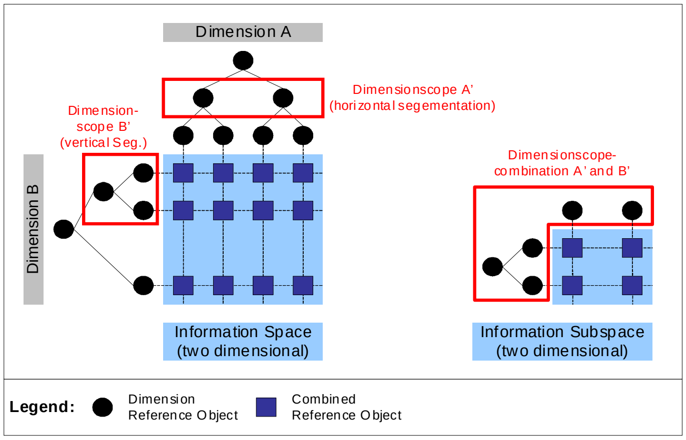
This concept allows a detailed description and analysis of web-based information systems like *asinfo*. The counterparts of the concepts in *asinfo* are as follows: Content refers to the documents indexed by *asinfo*, dimensions are the navigation trees (e. g. *Exposure Factors*), and reference objects are the elements (nodes and leafs) of each dimension (for instance Noise in *Exposure Factors*). All six dimensions (*Exposure Factors, Single Topics, Employment Protection Laws, Research, Economic Sectors, and Processes for Maintenance of Industrial Health and Safety Standards*) constitute the information space.

Exploration of the Conceptual Models

The navigational design of web-based information systems was investigated in research on conceptual modeling on hypertext and hypertext systems (for instance [Bot+92, Dhy+02, Niel90, Niel99]). In particular, web metrics were developed in order to describe these systems holistically and intersubjectively. Thus, there is a sound basis for analysis and comparison. We utilize and extend this research by applying it to the conceptualization of *asinfo*.

The information space of *asinfo* is spanned by the six dimensions shown directly on the homepage. Firstly, the dimensions are analyzed in terms of their complexity (e. g. number of reference objects) and their usage for document classification (assignments of content). The results varied greatly. For instance, the dimension *Exposure Factors* contains 185 reference objects (42 nodes, 143 leaves) whereas the dimension Research contains only the root node. Since each content is assigned to at least one reference object, it is interesting to investigate the colonization of reference objects with content (i. e. amount of documents that are assigned to each node). Moreover, the relative share of the total amount of reference objects for each dimension can be contrasted with the relative share of contents assigned to reference objects within one dimension. The result yields information about the distribution of content and reference objects in the information space. Furthermore, the variation coefficient of content for each dimension can be used to analyze the colonization of nodes in each dimension.

Figure 1. Information Space Conceptualization



The dimension *Exposure Factors* comprises nearly 75 % of all reference objects used in the system, but contains only 54 %of all content assignments (see Figure 2). The dimension Economic Sectors, on the other hand, contains 29 % of all content assignments but comprises only 7 % of all reference objects. On average, each node in the dimension *Economic Sectors* contains over 3000 documents, whereas the average in *Exposure Factors* is 554. Thus, information retrieval effectiveness for users may be lower in *Economic Sectors*. Additionally, the variation coefficient (content) is high in both dimensions, indicating that the colonization of nodes varies more than the other dimensions. Average content length in pages (41.59) and content length variance coefficient (1.45) characterize *asinfo* as a digital library or archive. On average, each content is assigned to two reference objects.

APPLICATION OF THE GOM FRAMEWORK TO THE *ASINFO* CASE

Framework for the Analysis

The navigational structure of *asinfo* can be interpreted as a model of the documents hierarchy that is contained in the repository, depicted by the *Model of* arrow (see Figure 3). The creators of *asinfo* designed this navigational structure for a certain purpose (distribution of information on industrial health) and specific addressees (users of *asinfo*). The purpose and addressees, both correspond to the constructs in the meta model of the GoM principles (see ERM in [Bec+95, p. 436]).

As observers of *asinfo*, we consider the system itself as a black box. The construction of the conceptual models is validated using the GoM framework (*Model Construction*). Moreover, we construct and apply a measurement framework to *asinfo* (*Model Application*). By calculating measures, we are able to condense information about *asinfo*. Moreover, we can assess *asinfo*'s navigational structure indirectly (*Adequacy of the Model*).

Figure 2. Selected Key Figures

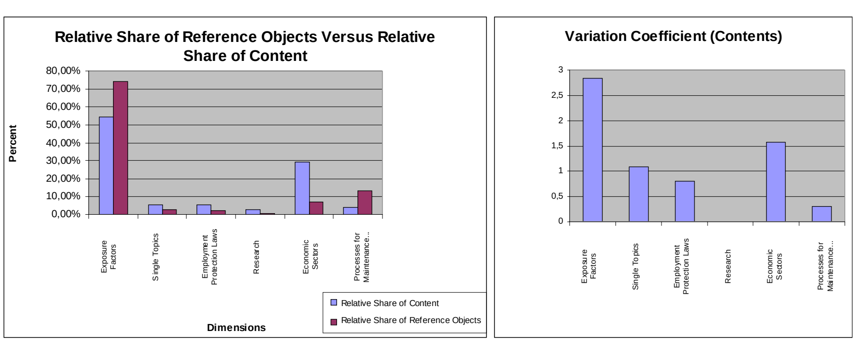
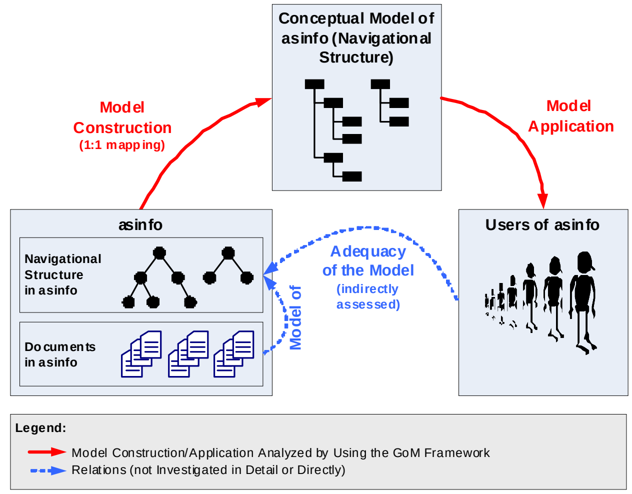


Figure 3. Analysis Framework



Application of the GoM to Models of *asinfo*

As depicted in Figure 3, the conceptual model is congruent with the *asinfo* website in all aspects to be explored. This construction is necessary in order to highlight certain facets of the original information already available from the *asinfo* website that can be depicted in this form more easily. With respect to the GoM principles, all semantic requirements (including the scope) are considered to be met, because we assume that the original data is semantically correct. The syntactical constraints the GoM impose, are complied with as detailed in the conceptualization described in Section 3.2. Syntactic comparability is only appropriate whenever different modeling techniques are used.

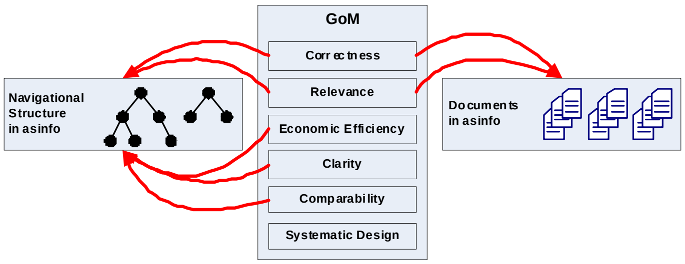
This leaves for further exploration, the Guideline of Relevance, in particular the level of abstraction, and the principles of economic efficiency and clarity.

Whereas the latter two are influenced solely by the procedure model and its underlying technique when creating the models (i. e. in terms of cost/time efficiency and ease-of-use of the conceptual model), the level of abstraction helps to establish the appropriate focus for the model application by choosing the relevant distinctness. In fact, the model may not capture all relevant objects of *asinfo*'s system, since not all aspects of the original data are incorporated. Abstractions made include the omission of the number of links on a single page, the link ordering, and further arrangement parameters.

More significantly, the model application offers insight into the model adequacy. The conceptual model as well as the information system itself can be assessed (indirectly)as shown in the following list:

- Semantical correctness implies that users find correct information and that the information is structured according to their needs. This means, for instance, that only current documents can be found in the six different dimensions and that no out-dated information is available and that these documents are indexed accurately.
- Furthermore when visiting *asinfo*, users not only expect correct information but also a certain range. The degree to which the information supply fits their expectations can be expressed by the relevance criterion. This means that *asinfo*'s offer of information is adequate if any document the user might hope to find in the context of industrial health, safety standards, and provisions on labor, is available. Moreover, the degree of abstraction used in our models influences this criterion. Whenever the level of abstraction is too high, either the model is inadequate or *asinfo*'s navigational structure is insufficient, since we certainly cannot offer a more detailed model than the original structure.

Figure 4. Influence of the GoM on asinfo



- Users of *asinfo* use the model to search *asinfo*'s repository for relevant documents, *asinfo* uses the platform to aggregate and communicate these documents. Thus, economic efficiency refers to the application of this model by both parties. The most efficient model of a navigational structure for users would result in a (global) minimum of search effort. Navigational structures for the main target groups of users (clusters) may therefore lead to the most efficient average search time.
- The Guideline of Clarity assures whether the model is readable, well structured, and clear. With respect to *asinfo*, this measure can be applied to the user's understanding of the navigational structure. Clarity of arrangement is assured, because of the common representation as a navigation tree. Readability refers to the naming of the different navigation levels. It is necessary to ensured that distinct names are chosen and repetitions in alternating levels within the navigation hierarchy are kept to a minimum. When repetitions are necessary, they must be on the same level.
- Comparability of the navigational structure of *asinfo* to other navigational structures of web-based information systems is a main criterion of the Guideline of Comparability. Since any user working with *asinfo* will be using other websites that may have other ways of guiding the user through the information offered, *asinfo* has to abide with the generally accepted methods of navigation in the context of web-based information systems.

Figure 4 shows the influence of the five relevant GoM on either the adequacy of the navigational structure or the selection of documents.

CONCLUSION

In our paper, we analyzed a web-based information system in order to provide operational means for its improvement. By integrating a sound conceptualization with formal measures and the GoM framework, we were able to derive meaningful and theoretically founded criteria for this purpose. These criteria can be instantiated in order to assess other web-based information systems with a particular focus on navigational issues. The main contribution of our approach is the integration of theoretically sound approaches and means that allow an intersubjective and comparable analysis of this particular type of information system.

Having correlated the GoM framework with web-based information systems, further work could concentrate on providing operational criteria for some of the GoM by using the measurement framework proposed in Section 3.3. Depending on the information system (its purpose and scope), its type (e. g. commercial, intranet, archive) and its users (experts versus non-experts) different recommendations may apply. However, the measures can be used to perform and underpin such an in-depth analysis. Moreover, more complex navigational structures (especially personalized and different types of navigation) can be assessed through an extended framework.

REFERENCES

[Bat+92] Batini, C.; Ceri, S.; Navathe, S. B.: Conceptual Database Design. An Entity-Relationship Approach. Redwood City, California, 1992.

[Bec+03a] Becker, J.; Brelage, C.; Klose, K.; Thygs, M.: Conceptual modeling of semantic navigation structures. The MoSeNa-approach. In: Proceedings of the 5th ACM international workshop on Web information and data management (WIDM 2003). New Orleans, 2003, pp. 118-125.

[Bec+03b] Becker, J.; Dreiling, A.; Holten, R.; Ribbert, M.: Specifying Information Systems for Business Process Integration - A Management Perspective. Information Systems and e-Business Management 1(3), 2003: pp. 231-263.

[Bec+03c] Becker, J.; Brelage, C.; Thygs, M.; Ribbert, M.: Conceptual Design of WWW-Based Information Systems. In: Proceedings of the 11th European Conference on Information Systems (ECIS 2003). Naples, 2003, pp. 118-125.

[Bec+95] Becker, J.; Rosemann, M.; Schütte, R.: Grundsätze ordnungsmäßiger Modellierung. Wirtschaftsinformatik 37(5), 1995: pp. 435-445.

[Bot+92] Botafogo, R. A.; Rivlin, E.; Shneiderman, B.: Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. ACM Transactions on Information Systems 10(2), 1992: pp. 142-180.

[Dhy+02] Dhyani, D.; Ng, W. K.; Bhowmick, S. S.: A Survey of Web Metrics. ACM Computing Surveys (CSUR) 34(4), 2002: pp. 469-503.

[FaFa00] Farkas, D. K.; Farkas, J. B.: Guidelines for Designing Web Navigation. Technical Communication 47(3), 2000: pp. 341-358.

[Hev+04] Hevner, A. R.; March, S. T.; Jinsoo, P.; Ram, S.: Design Science in Information System Research. MIS Quarterly 28(1), 2004: pp. 75-105.

[Holt03] Holten, R.: Specification of Management Views in Information Warehouse Projects. Information Systems 28(7), 2003: pp. 709-751.

[LaCz98] Larson, K.; Czerwinski, M.: Web page design: implications of memory, structure and scent for information retrieval. In: Proceedings of the Proceedings of the SIGCHI conference on Human factors in computing systems. 1998, pp. 25-32.

[Lin+94] Lindland, O. I.; Sindre, G.; Sølvberg, A.: Understanding Quality in Conceptual Modeling. IEEE Software 11(2), 1994: pp. 42-49.

[Maie96] Maier, R.: Qualität von Datenmodellen. Wiesbaden, 1996.

[MoSh03] Moody, D. L.; Shanks, G. G.: Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework. Information Systems(28), 2003: pp. 619-650.

[MoSh94] Moody, D. L.; Shanks, G. G.: What Makes a Good Data Model? Evaluating the Quality of Entity Relationship Models. In: Proceedings of the 13th International Conference on the Entity-Relationship Approach - ER '94. Business Modelling and Re-Engineering. Manchester, 1994, pp. 94-111.

[Niel90] Nielsen, J.: The Art of Navigating Through Hypertext. Communications of ACM 33(3), 1990: pp. 296-310.

[Niel99] Nielsen, J.: User Interface Directions for the Web. Communications of ACM 42(1), 1999: pp. 65-72.

[Sha+03] Shanks, G. G.; Tansley, E.; Weber, R.: Using Ontology to Validate Conceptual Models. Comm. of the ACM 46(10), 2003: pp. 85-89.

[Wan+99] Wand, Y.; Storey, V. C.; Weber, R.: An Ontological Analysis of the Relationship Construct in Conceptual Modeling. ACM Transactions on Database Systems 24(4), 1999: pp. 494-528.

Related Content