Photogrammetrie • Fernerkundung • Geoinformation 3/2008, S. 217–225, 5 figs.

Integration of Language in GIS: Models in Ownership Cadastre and Disaster Management

CHRISTIAN LUCAS, MARINA MUELLER & HANS-PETER BÄHR, Karlsruhe

Keywords: Disaster Management, Data Modeling, Abstraction, Knowledge Representation

Summary: This paper discusses relevant aspects of interrelationships between graphically and verbally represented information. The research is focused on both ways of transformation – from textual to graphical information and from graphical to textual information – with respect to automation. Analyses based on examples of *Boundary Descriptions of the Brazilian Ownership Cadastre* as well as *Communication Systems in Disaster Management* are presented. Due to the various levels of detail of given attributes, the textual representation contains various abstraction levels. Such experiences are implemented in terms of data modeling and processing.

Zusammenfassung: *Integration von Sprache in GIS*: Modelle im *Eigentumskataster und Katastrophenmanagement*. Dieser Beitrag präsentiert theoretische Untersuchungen zur Verarbeitung graphischer und sprachlicher Zusammenhänge. Den praktischen Hintergrund dieser theoretischen Ansätze lieferten zum einen die verbalen *Grenzbeschreibungen des brasilianischen Katasters* und zum anderen die verbal ausgerichteten *Kommunikationssysteme im Katastrophenmanagement*. Dabei zeigte sich, dass die textuell repräsentierte Information auf Grund der verschiedenen Detaillierungsgrade der Beschreibung verschiedene Abstraktionslevel bedingt. Dies wurde speziell bei der Modellierung der Wissensbasis berücksichtigt.

1 Background

1.1 *The Necessary Step from "Real World" to a Symbolic Level*

Language in written or spoken form is certainly the most important medium for human communication. Therefore, language represents a very important factor in GIS, too. The formal neighborhood of language to the GIS domain and to image analysis is proven by many metaphors taken from linguistics and transformed to geospatial descriptions. These are for instance *context, understanding, redundancy, completeness*, or *abstraction*. The listed examples, which may be completed, reveal a semantic neighborhood of verbal and graphical descriptions of our environment. The presented paper does not give a contribution for automatic language recognition but aims to show how to integrate verbal or written language into formal spatial models. This task is basically of the same nature as data fusion. Textual processing together with images or maps requires as a necessary first step the transformation of data from the iconic (or *verbal*) level to a symbolic level.

This scenario is formally presented in Fig. 1. Images, maps and text to be fused are taken from a "real world" which cannot be captured directly by human cognition. However, its transformation yields images, maps or texts, according to specific tools applied. On the iconic or verbal level which is directly accessible by humans, numeric processing or fusion is impossible. In order to do so, a second transformation to the so-called symbolic level is necessary which ends up with a formal representation of the three

1432-8364/08/2008/217 $ 2.25

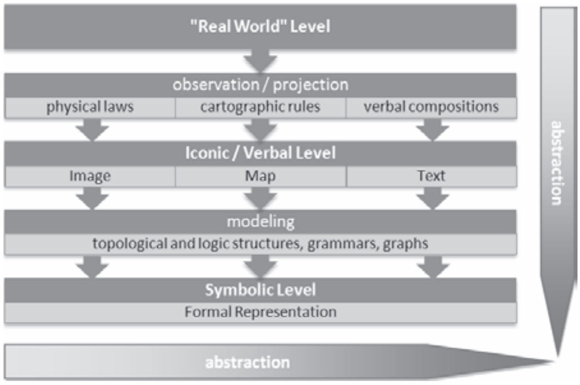

Fig. 1: Levels of knowledge representation from the ''real world'' to a formal representation.

mentioned components image, map and text. The step from iconic to symbolic is called *modeling* and may be performed by very different tools like topological and logical structures, grammar and graphs. The result is a closed formal knowledge representation at a high level of abstraction, an ontology[1], respectively.

Both steps, from real world to formal representation as well as from image to text increase the level of abstraction. In consequence images are closer to reality than texts and pictorial, cartographic or textual representations are closer to human perception than e. g., a semantic network used for modeling. Therefore, integration of text into geospatial data sets necessarily includes components of data abstraction.

1.2 *Introduction of two Applications*

A straightforward application where the intimate interrelation of language and maps can be observed is way finding or navigation. This has already been widely recognized (Tversky 2003). Another topic related to this application is the description of borders used in ownership cadastres, where borders of parcels may be considered as a navigational task, i. e., a path to be followed. Thus the research is focused on the *Brazilian ownership cadastre*. Legal validation of the ownership in Brazil is primarily verbal and not based on coordinates or a similar graphic representation, however the available data consisted of descriptions and accompanying maps. The verification of consistency between both levels, graphic and verbal, gives a sound example for combining graphs (maps) and language (text) together in the context of practice.

The research topic *Communication Systems in Disaster Management* focuses on transformations of verbal descriptions in German language concerning the status and location of disaster events. Aim is to represent such information automatically and up to date by a GIS. During disaster events, the emergency operations center (EOC) receives hundreds of textual reports from damage sites, given by several on-site units and passer-by. The information of all reports needs to be evaluated as soon as possible by an operator of the EOC and added into a situation map that serves for information sharing as well as a basis for decision making.

2 The Abstraction Levels

2.1 *Abstraction Levels of Textual Descriptions in the Brazilian Ownership Cadastre*

Current texts in the Brazilian rural ownership cadastre are formulated very precisely in general. Such accurate texts are generated

[1] The term ontology is defined as an explicit specification of a conceptualization (Gruber 1993).

by surveys where coordinates of a starting point and the complete polygon of the boundary with length and angle of direction for each line are measured as exactly as possible. Due to missing standards for texts of the Brazilian land register up to the year 2001, levels of detail and styles of older descriptions vary strongly depending on the author of the cadastral text. These texts cannot be assigned to a single level of abstraction, as assumed at the beginning of the research. Various geometrical attributes (e. g., distance *and* direction) of *one* boundary as well as a single attribute (e. g., distance *or* direction) of *some* boundaries correspond to different levels of detail as well as different levels of abstraction. Within the explicit knowledge representation (here an ontology) the abstraction hierarchies are integrated so that all occurring combinations of geometric attributes are considered. Due to the intended transformation from textual to graphical representations, the abstraction hierarchy is based on transformation aspects of necessity and uncertainty. Thus the transferable information always corresponds to one of the four derived hierarchical levels of geometric attributes:

- none/general knowledge: All necessary information for interpreting a single boundary attribute is readily available by the text passage. Example: *It begins at point 01 with the geographical coordinates N = 7.734.679,703 and E = 248.328,107.*
- context dependent knowledge: Information which is given for a single attribute cannot be reconstructed independently because of included references to already given information within the textual description. Example: the direction of the previous edge is essential to handle phrases like *from here it turns left for 30°*.
- external sources: Information include references to objects or spatial attributes which has to be acquired from external sources. Example: the course of the river is essential to handle phrases like *it runs along the river for a length of 162 m.*
- all/new acquisition: Usable information is completely missing and has to be acquired on site. Example: *the area extends in depth as far as the cows graze.*

More detailed explanations for the Brazilian land register as well as the abstraction levels appearing in texts are in (Mueller 2008).

2.2 *Abstraction Levels of Reports in the Domain of Disaster Management*

The format of reports in the domain of disaster management is different to the descriptions of the Brazilian land register. Disaster reports, given by on-site units and passer-by, consist of fixed entries, e. g., for time stamps, sender and receiver, as well as a free-form text element where current information can be given in natural language. The report templates usually contain for a representation in a situation map both, information about relevant facts like the damage sites including its spatial references and also information about irrelevant facts for visualization like the resources, which has to be separated. This is in contrast to the cadastral descriptions were all the given information are relevant. The level of detail of information and thus the level of abstraction vary as well because of the different reporting persons. Consequently the subdivisions of the main abstraction levels for a disaster management application were adopted from the Brazilian cadastre application. However the characteristic of the levels of abstraction is adjusted as follows:

- none/general knowledge: All necessary information for interpretation of a single fact is given by the report. This case is possible but implausible because practically no-one uses coordinates for reporting a disaster event.
- context dependent knowledge: The given fact cannot be interpreted independently because it refers to previous reports. Example: to handle the reported fact *at the accident are 3 injured people*, information about the existent accident is essential.
- external sources: The reported fact include references to objects or attributes which

have to be required from external sources. Example: to handle the reported fact *midland school is burning*, the location and function of buildings, here the midland school, are essential.
- all/new acquisition: The reported fact is not processible. Example: *a few minutes ago, I have heard something anywhere*. Acquisition on-site is usually difficult because of missing references.

A basic difference to the cadastral application is the "segmentation" of facts from a report as well as the autonomous processing of facts. This is already a result of the separation of relevant facts, but also a simplification. In consequence of this approach a single fact occupies only a single abstraction level. The successive arriving reports usually deal with different facts from different disaster sites. Nevertheless there are necessary references to already given reports, which have to be considered.

3 Semantic Augmentation and Knowledge Representation

3.1 *Common Symbolic Representation*

There are a number of tools supporting an explicit modeling of domain knowledge such as rule based systems or semantic networks. Different systems are often equivalent concerning their abilities to model different types of information and to draw conclusions from it. Therefore, the decision for one or the other tool depends on e. g., the form of explicit knowledge representation that is preferred (rules, networks) and the kind of tools that the knowledge engineer is familiar with (Sowa 2000). In the current application, a semantic network is chosen. Fig. 2 shows an example excerpt of a possible semantic network for the cadastre and disaster scenario that can be realized e. g., by tools such as Protégé (Protégé 2007).

Since both data sources, texts and maps, are supposed to include similar spatial and attribute information, they can be modeled with similar structures on the level that is displayed in Fig. 2. However, explicit geometric information is more relevant in a map representation than in a textual description. First differences become already visible at the displayed level of detail on closer examination of the data model. Each object node is modeled according to the frame-model by Minsky (1975) using a frame with different slots, i. e., characteristic attributes for each object. Fig. 3 shows cadastral examples for the frames boundary-part-in-text and boundary-part-in-map. Besides information that is identical for both frames (neighbor), other attributes such as the length of a boundary include different semantics for each source of information. The text refers to the length value of boundary parts which can be measured in a field survey. In the case of the map, this length value is included in written

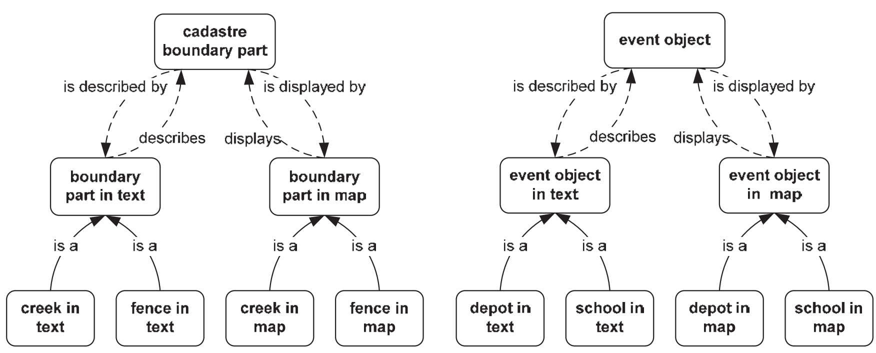

Fig. 2: Excerpt of a semantic network for the cadastre example (left) and a similar disaster example (right).

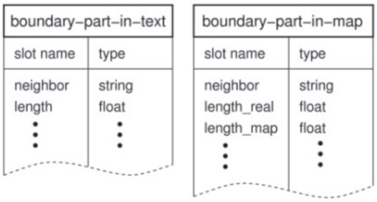

Fig. 3: Boundary-part-in-text and boundary-part-in-map of the semantic network.

numbers close to the drawn line. Additionally, this drawn line has its own length resulting from the scale of the map and from the precision in drawing.

3.2 *Knowledge Representation for an Application of the Brazilian Ownership Cadastre*

Basis for a transformation from texts into maps is the knowledge how textual phrases about geometrical information have to be interpreted in terms of discrete values in the context of the Brazilian ownership cadastre. While e. g., directional information like *an azimuth of 328.2°* can be directly used for map drawing, phrases like *turn slightly left* will give an interval of possible degree values. The ontology that describes how geometrical information is represented in texts of the Brazilian cadastre naturally contains the basic geometrical elements (points, lines with direction and length) and their typical relations such as *a line consists of two end points*. However in order to capture also the uncertainty that is connected with the texts, indicators of uncertainty – e. g., the use of quantifiers such as *slightly* – needed to be identified in the text corpus and integrated (including a quantification of the qualitative expression) at the associated level of abstraction of the affected geometrical attribute.

3.3 *Knowledge Representation for Communication Systems of the Disaster Management*

For the specific representation of knowledge existing data models offer a good starting point for developing ontologies. Such standardizations for the disaster management domain are given by the Emergency Data Exchange Language (EDXL) and the Common Alerting Protocol (CAP). Unfortunately, these data models are primarily oriented on the management of disaster resources. A detailed comparison of the diverse data models are given by Werder et al. (2006). It turned out that the NATO standard for military interoperability, the *Command and Control Information Exchange Data Model* (*C2IEDM*), satisfies in most instances the requirements of disaster management and provides a good basis for the development of the *Disaster Management Data Model* (*DM*^{2}). Because of focusing on information storage, retrieval and processing, there are a couple of differences between DM^2 and C2IEDM. Therefore, one of the most obvious changes is the philosophy of object representation. The focus of the domain specific ontology is on modeling the spatial aspects of objects. In this manner the common spatial attributes of objects are the location as well as geometric attributes of form, size and feature alignment. These elementary attributes, traditionally supported by spatial ontologies and geographical information systems (GIS), describe discrete objects unambiguously by their dimension and location in space. In order to support a spatial reasoning process for disaster events based on textual descriptions, a more comprehensive level of spatial information is necessary. The method of object modeling within the ontology has to be similar to the mental model of the reporting person (Frank 1998). This mental model contains, besides discrete objects, spatial scenes with interactions of two or more spatial objects in the meaning of neighborhoods or part-of-relations like include, overlaps or tangent. Such detailed information level is essential for analyzing e. g., coherences of cities, districts, damage sites and operation areas. Furthermore, time is modeled explicitly in the DM^2 as well as the mutable attributes of the objects. Mutable attributes in terms of object modeling are state and location. Such dynamic aspects are modeled by asso-

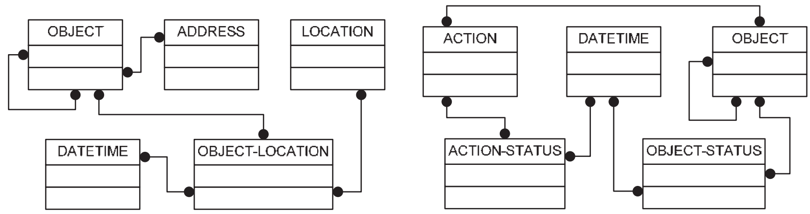

Fig. 4: Associations of Object, Location and DateTime (left) as well as the associations of Action, DateTime and Object (right) – in IDEF1X notation.

ciations of the relevant classes (cf. Fig. 4). In the DM^2 this was done by creating tuple of *object-state-time* and *object-location-time*. The concept as shown in Fig. 4 is explained in more detail in (LUCAS et al. 2007).

4 Results

4.1 *Results of the Application of the Brazilian Ownership Cadastre*

Generating Maps from Texts: Assuming a precise text that contains (a) a starting point with precise coordinates and (b) boundaries as straight lines with precise direction and length, a transformation to a map representation is simple. Necessary additional definitions for such a transformation are the desired *scale of the map, orientation of the map and local map coordinates of the starting point*.

Conversion of coordinates and length information of the text into map coordinates can be done by using straightforward mathematical computations. Saving the generated map information in widely used standards such as XML (Extensible Markup Language) allows its graphical visualization with different tools (e. g., ArcGIS from ESRI). Fig. 5 shows an example of an automatically generated map and its corresponding original map. Although the general outline of the boundaries is identical, the generation of the new map reveals that the north arrow of this otherwise very precise original map deviates about 15 degrees from the true northward direction. Such variations of the north arrow within maps of the Brazilian land register are not unusual because of inaccuracies of observation.

If the text is vague, with unknown references and missing essential boundary information, the generation of maps usually requires additional sources of information (maps of neighboring real estates, topographical maps, or remote sensing data such

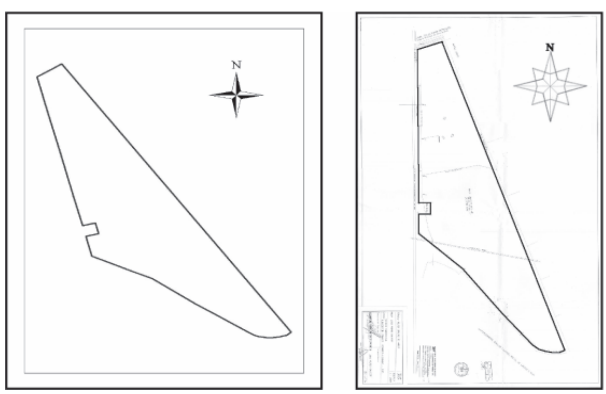

Fig. 5: Automatically generated map from a precise text (left) and the cadastral map (right).

as aerial images). However, vague sketch maps are often useful to give a first overview of the actual information provided in texts and even vague texts include constraints, such as

- length of each boundary part or
- end point of the description = starting point of the description.

Additional to such constraints, a reasonable quantification of qualitative expressions (e. g., left / right) and further heuristics (such as typical errors in descriptions) can be used to generate maps even from texts with incomplete information (Mueller 2008).

Generating Texts from Maps: Besides geometrical information given as drawn lines, boundary parts in maps of real estates include further attributes. Such attributes are the type of the boundary part displayed by the special type of line drawing as well as directly given information like length and direction of the real boundary part as written text in the map.

Given a precise map with detailed attribute values, the generation of a text only requires the definition of (a) a starting point and (b) the direction of description (clockwise/ counterclockwise). Topological information such as connectivity between boundary parts can be modeled in the semantic network. This information can be directly used to deduce the succession of boundary parts in the description.

In case a map is only a sketch of the real situation, such a vague map can nevertheless be transferred to a vague text where relative relations between boundary parts are expressed by relative textual information. In order to generate a text from a vague map, a starting point has to be chosen and to be described as meeting point between neighboring real estates. Similar to the generation of a precise map, the direction of the description (clockwise / counterclockwise) has to be decided.

4.2 *Results of the Application of the Disaster Management Domain*

For processing, a computational representation of the reported information is needed. A standard formalism of computational linguistics for representing extracted information is the *typed feature structure*. These typed feature structure is an XML-representation (Extensible Markup Language) of the relevant information. Here, the extraction was done by an operator, who has to register the relevant information into the graphical user interface (GUI, an input mask for the domain requirements) and the application generates the XML.

Reports of the disaster management domain in general include temporal uncertainties as well as semantic and spatial ambiguities. By explicit modeling of temporal aspects in the DM² (cf. Section 3.3), a basis for both, historiography and temporal reasoning are given. Temporal reasoning provides a possibility for solving ambiguities by using interval algebra[2] and allows building up a temporal context. This temporal context is focused on searching for events which take place in the same frame in time. Detailed considerations of domain specific temporal logic are given by Werder et al. (2007). Semantic ambiguities are amongst others a result of alternative (*Fußballplatz* vs. *Bolzplatz*), multiple (*Bahnhofstraße*, exists 3 times) and abstract (*church* without specific name) object denotations and can be limited or solved by a-priori knowledge, given by the DM². In consequence of the diversity of this task, collecting as well as modeling of such background knowledge is very comprehensive and complex. Spatial ambiguities are an effect of using adverbial phrases and spatial identifiers. Such terms like *behind* or *north* of are already given by descriptions of spatial scenes within the reports. Because of the diverse types of ambiguities different approaches for reducing and solving are necessary.

[2] Here the interval algebra of (Allen & Ferguson 1994) is used.

One method for solving is using context knowledge based on the reports. Initial situation for processing is a range of reports describing one or maybe more events[3]. First indicator of dependence between diverse reports and their facts is the time frame of receiving. Thus the first context is a temporal one. Next point of interest for processing and basis for creating a further level of context is the location. If compatible events like a *cloud of gas* and an *explosion* are at the same frame in time and close to each other, coherence is probable. Proximity for creating a local context is possible by the introduction of the so-called event horizon, which defines the sphere of influence of a specific event. This sphere of influence depends in shape and size on the reported fact. According to that approach, the reported facts are related whenever their event horizons overlap. This location context now offers possibilities to analyze and evaluate the reported facts. In this knowledge based reasoning process the reported facts are first evidences for a respective event, a so-called hypothesis. The reasoning is made possible by defining conditions and relations between them. These basic relations are $A \wedge B$, $A \vee B$ and $A \rightarrow B$ which are quite simple relations but adequate for representing the important dependences. Such dependences are for example, *gas* $\wedge$ *explosion* $\rightarrow$ *fire*. Furthermore the relations are complemented by a certainty factor which represents a weighting for evaluation. That way the credibility of a fact respectively the credibility of the reporter can also be considered. Result of this reasoning is an array of possible events including an evaluation. For the above-mentioned example of the gas and the explosion, the possible hypotheses are the events *fire*, *toxic cloud* and *building damage*. Detailed explanations of this procedure are given by (Lucas et al. 2007).

[3] According to definition, an event relates to just one object.

5 Conclusions and Outlook

Due to the various levels of detail for spatial attributes within descriptions of spatial scenes, four levels of abstraction were defined for an application of the Brazilian ownership cadastre. These levels allow processing of spatial phrases under considerations of the inherent uncertainty. For processing, an adequate representation of knowledge is also necessary, especially for including context dependent as well as external knowledge. For a disaster management application, the levels of abstraction were adapted from the cadastral application and adjusted to the domain requirements. The verbal representation of the facts demands the use of complex structures of knowledge modeling and representation, as given by ontologies. For a more robust spatial reasoning process the linguistic vagueness and the mental model have also to be taken into account. Furthermore the area of validity for the spatial identifiers has to be represented exactly by a weighting function e. g., by fuzzy logic.

Acknowledgements

The research was sponsored by the Deutsche Forschungsgemeinschaft (DFG), project no. BA 686/16. Moreover, this work was supported by Stefan Werder (Collaborative Research Center 461).

References

Allen, JF. & Ferguson, G., 1994: Actions and Events in Interval Temporal Logic. Journal of Logic and Computation. Special Issue on Actions and Processes **4** (5): 531–579.

Frank, A., 1998: Formal models for cognition – taxonomy of spatial location description and frames of reference. – Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, LNAI 1404, Springer: 293–312.

Gruber, T-R., 1993: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, Volume **5** (2): 199–220.

Lucas, C., Werder, S. & Bähr, H.-P., 2007: Information Mining for Disaster Management.

– International Archives of Photogrammetry, Remote Sensing and Spatial Information Science **36** (3): 75–80.

Minsky, M., 1975: A framework for representing knowledge. – The Psychology of Computer Vision, McGraw-Hill, New York: 211–277.

Mueller, M., 2008: Transformations of Cadastral Descriptions with Incomplete Information into Maps. – Transactions in GIS **12** (1): 83–101.

Protégé, 2007: Homepage, protege.stanford.edu (last visited: December 2007).

Sowa, J.-F., 2000: Knowledge Representation: Logical, Philosophical and Computational Foundations. – Pacific Grove, Brooks/Cole, CA.

Tversky, B., 2003: Navigating by mind and by body. – Spatial Cognition III: Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Reasoning, Springer: 1–10.

Werder, S., Mueller, M., Mueller, M. & Kämpf, C., 2006: Integrating Message Information into Disaster Management Maps: Transferability of a System of the Military Domain. – ISPRS Symposium on Geospatial Databases for Sustainable Development, Goa, India, on CD.

Werder, S., Lucas, C. & Bähr, H.-P., 2007: Information Extraction from Messages in Disaster Management. ISPRS Symposium on Geoinformation for Disaster Management, Toronto, Canada, on CD.

Addresses of the Authors:

Dipl.-Ing. Christian Lucas, Prof. Dr.-Ing. habil. Dr. h.c. Hans-Peter Bähr, Universität Karlsruhe, Institut für Photogrammetrie und Fernerkundung, Kaiserstrasse 12, 76128 Karlsruhe, Tel.: +49-721-608-2315, Fax.: +49-721-6088450, e-mail: christian.lucas | hans-peter.baehr@ipf.uni-karlsruhe.de

Dr.-Ing. Marina Mueller, COMSOFT GmbH, Wachhausstrasse 5a, 76227 Karlsruhe, Tel.: +49-721-79076-84, Fax.: +49-721-79076-89, e-mail: marina.mueller@comsoft.de

Manuskript eingereicht: Dezember 2007
Angenommen: März 2008