ON THE ACQUISITION, REPRESENTATION AND APPLICATION OF KNOWLEDGE IN GEO-INFORMATION SYSTEMS

Ralf Bill, Senior Researcher, Institute of Photogrammetry, University of Stuttgart, Federal Republic of Germany, Commission III

30th April 1992

ABSTRACT

The paper presents the concepts of knowledge acquisition, representation and application with respect to their usage in geo-information systems (GIS). Recent database research activities are partially integrated into commercially available GIS-products whereas the use of knowledge administration and processing on the other hand is far away from any development in the GIS-community. Examples should illustrate their general applicability in the GIS-area.

Beginning with some definitions on data, information and knowledge various forms of knowledge are mentioned. The knowledge base is the major part of a knowledge-based GIS; there the rules of type 'If - then' are stored. In GIS-products one can find knowledge about rule-based relations and prototypical knowledge integrated in a procedural manner. The integration of fuzzy knowledge would be usefull for certain applications.

The *knowledge representation formats* are treated in detail, they are logic models, production rules, semantical networks and frames. Similarities to database research activities such as the entity-relationship-model and object-oriented are discussed.

The *architecture* of a GIS differs very much from the demand of a knowledge-based system. The procedural design of a GIS needs to be replaced with the inference-mechanism of an expert system, heuristical approaches instead of procedural ones are to be integrated.

Different GIS-applications of *knowledge integration* such as fuzzy-polygon overlay, rule-based raster-vector conversion, object modelling based on semantic networks or frames are presented and the advantages illustrated.

KEY WORDS : Knowledge acquisition, knowledge representation, knowledge-based systems, GIS

1 INTRODUCTION

1.1 Definitions

At the beginning of this paper the major terms 'data, information and knowledge' should be defined. It is not the intention of the author to define these in an overall context, because this would end in a philosophical discussion. So, the given definitions are context-related. Figure 1 tries to illustrate the relation between data, information and knowledge in a simplified manner.

Data in the classical meaning are simply characters. These data can be read, stored, compared, processed and written by a computer. Extending the definition of character data could also be images, texts, language, too. Data itself are of meaning for a computer but without any knowledge about the meaning and structure of the data a human being is not able to read and interprete it easily.

Information is related to data, but we want to see it here as a result of using transformations, rules and knowledge familiar to those working with the data to gain facts and interpretable results in a given context. Knowing the way data are organized in a computer and the purpose of the data storage in the computer a human being is able to work with the data. Information is to be differentiated in three levels; the syntax, i.e. the characters used, the semantics, i.e. the meaning, and the communication, i.e. the way information is distributed.

The *knowledge* of a knowledge carrier is defined as the sum of all declarations about the represented part of the world regarded as true, that are really true. *Persuasion* in comparison to knowledge is all what the knowledge carrier believes to be true.

1.2 Spatial information systems

Spatial information systems (or geographical or geo-information systems (GIS) as synonyms) are computer-based systems which consists of hardware, software, data, and the applications. A spatial information systems supports the digital input, management, analysis, and presentation of spatial data (R. Bill and D. Fritsch (1991), P. A. Burrough (1986)). Spatial or Geo-Information systems are combining the database technology with the information system technology under the common aspect of treating spatial phenomena. For both parts of technology this is not to be seen as a standard problem. In computer science only a small group of researchers is currently working on spatial problems. This led to the situation we have today, that a large gap exists between the research domain in computer science and the development and application area of GIS. Only a small number of GIS-products currently available on the market is close to the research activities e.g. in object-oriented databases. The situation becomes even worse if we look at the topic of this paper, the integration of knowledge in a GIS.

Data in a GIS Currently available GIS-products are dealing with the following data types :

- geometry : the coordinates and their metric. Geometric primitives are points, lines and surfaces in 2 or maximum 2.5 dimensional way.
- topology : the relation of geometric primitives and objects in itself and to others. Topologic primitives are nodes, edges and meshs. Their relations are treated in sharing, connectivity, neighbourhood, inclusion etc.
- attributes : information describing the non-geometric properties of the objects.

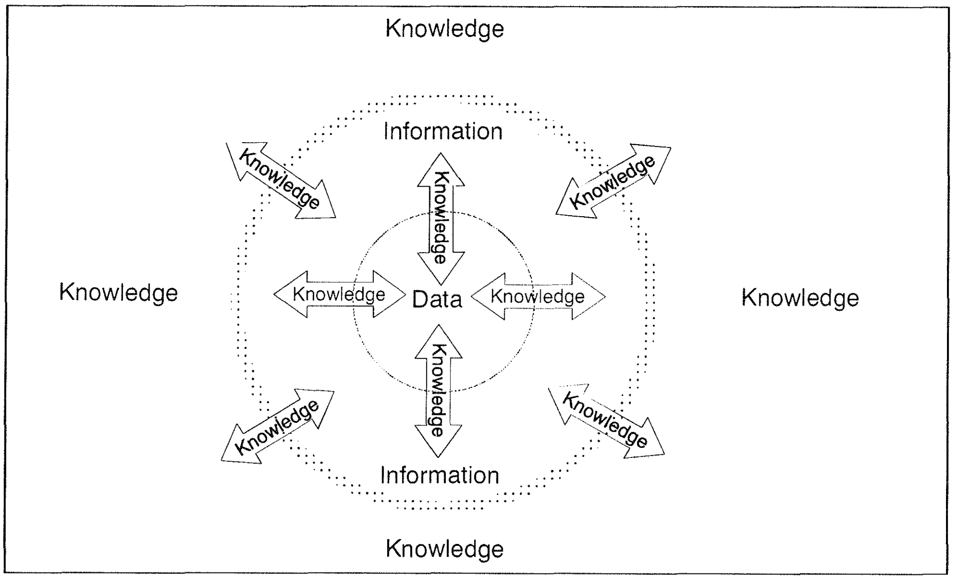

Figure 1 : Relation between data, information and knowledge.

- graphic transforms : information describing the graphical representation of the objects under a certain theme. These are symbol types, line styles, fill modes etc.

The data in a GIS are represented in the following logical data models (R. Bill and D. Fritsch, 1991). Hierarchical or CODASYL models are often used to store and maintain the geometry and topology in a GIS. Relational data models, the standard in commercial databases, are more and more used to handle the attribute data in a GIS. In some cases also the geometry and topology is stored and treated in relational form. Object-oriented data models is one of the most often used keywords in GIS, which is in most cases misused if one follows the definition of computer scientists. Object-oriented data models will be of importance in the future for GIS. Today there are only a limited number of products available on the market which are partially object-oriented. The term 'object-oriented' should, therefore, be illustrated in more detail following the computer-sciences (for further references see R. Bill and D. Fritsch (1992)). They are separating three types of object-oriented data models : the structural, the behavioural and the fully object-oriented systems mainly differing with respect to flexibility. Structural object-oriented systems are able to handle complex objects as an atom in a data base having predefined operators for these data types. Behavioural object-oriented systems support the definition of application specific data types and operators, so called methods. Fully object-oriented systems are combining the properties of both types of data models. An object consists of data and methods. The object itself is a small world in its own (called encapsulation). The object structure and behaviour may only be manipulated with the methods belonging to the object. A message is send via a channel to manipulate the encapsulated object which is illustrated in figure 2. Object-oriented systems are having close relations to some concepts of knowledge representation, which we will discuss afterwards.

2 KNOWLEDGE INTEGRATION IN GIS

2.1 Types of knowledge

Various types of knowledge are illustrated in figure 3, which should not be seen as disjunct sets but as correlative forms of knowledge. In addition the figure tries to illustrate what types of knowledge are implicitely or explicitely describable in a GIS. In particular the implicit form of knowledge is written down in the data model and the procedural flow of the GIS. GIS related examples for the various form of knowledge are given in detail in R. Bill (1991). Beside the contradictionary knowledge all other forms of knowledge could (and should) be integrated in a GIS. Today the assumption of completeness and exactness in the GIS-data model and the GIS-procedures are neglecting all types of uncertain and incomplete knowledge.

2.2 Architecture of knowledge-based systems

In a knowledge-based system the knowledge base is the major part of the system; there the rules of type 'If - then' are stored. In GIS-products one can find knowledge about rule-based relations and prototypical knowledge integrated in a procedural manner. The integration of fuzzy knowledge would be usefull for certain applications. Figure 4 shows the design of a knowledge-based system compared to a conventional program system: the last may stand for the typical GIS-product. One can easily see the difference lying in the flexibility of the problem solving strategy and in the domain specific knowledge which are integrated parts of a knowledge-based system. This would make it much more flexible to design an application-specific GIS. But the problem today is that the architecture of GIS-products is not very flexible. GIS today completely belong to the group of conventional programming systems. This leads to the current situation of knowledge integration in GIS, where the knowledge part is completely done outside the GIS. The GIS is seen as a data capture and maintainance tool with the ability of visualizing results. To make use of the development of knowledge-based systems the procedural design of a GIS needs to be replaced with the inference-mechanism of an expert system, heuristical problem solving approaches instead of fix-coded procedural ones are to be integrated.

2.3 Knowledge acquisition

Usually in an application domain, which intends to make use of GIS technology, the whole expertise i.e. the knowledge about the application is with the application specialists. Thus, the carrier of the knowledge is not the GIS developper and, in some cases also, not the GIS user. The problem is to transfer the knowledge of the application specialists to the one who is responsible for

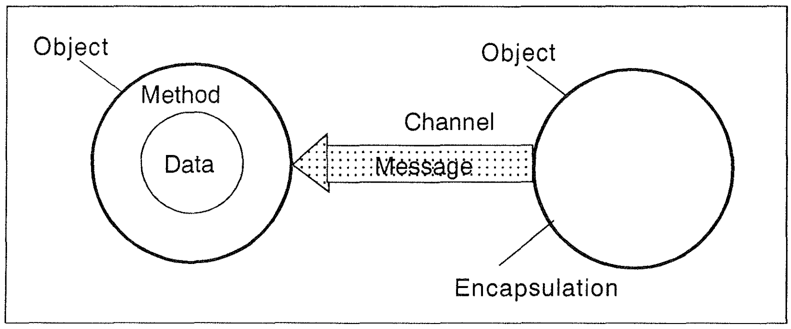

Figure 2 : The object-oriented approach.

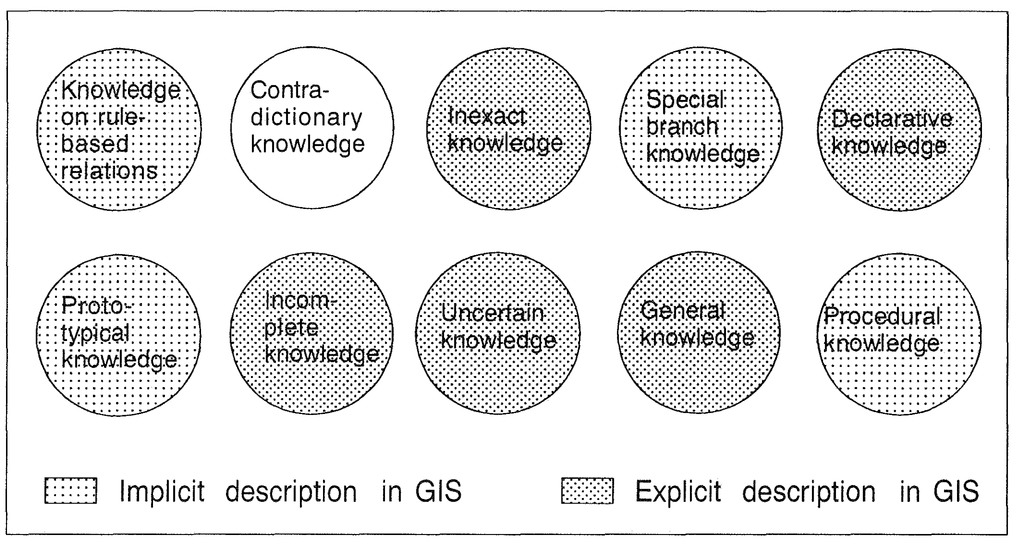

Figure 3 : Various types of knowledge.

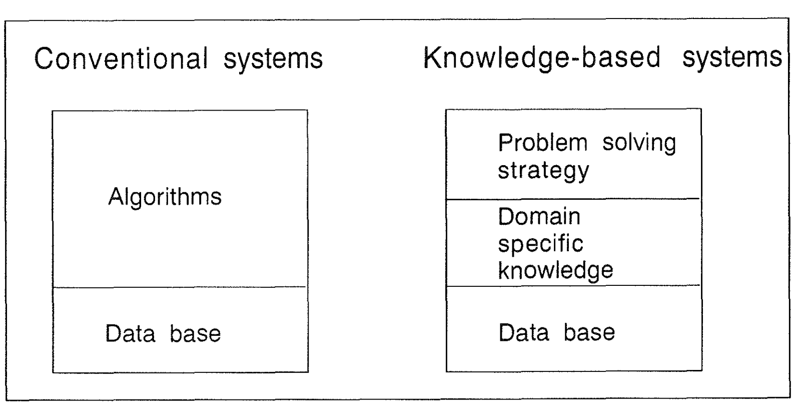

Figure 4 : Conventional versus knowledge-based systems architecture.

tuning the GIS to the application needs. Only the knowledge which is transferred to the developper will be available in the system later. Knowledge acquisition, however, is not only the transfer of expertise. Knowledge acquisition is a creative process in which the systems builder constructs qualitative models of human behaviour. This makes it necessary to find a structured approach for knowledge acquisition. In artificial intelligence well known methods of knowledge acquisition are :

- interviews such as interviews in standardised form or interviews focussed on a specific problem,
- group discussions, where several specialists are discussing about problem solving,
- observations and own experience, when joining the application domain,
- contents analysis, which includes study of literature as well as study of the manuscripts of interviews and group discussions,
- learning methods (cf. E. Rich (1988)) such as routine learning, conceptual learning, learning by analogy, learning in a general problem solver, learning by exploring etc, which are often used for machine learning purposes.
- others.

From those methods especially conceptual learning seems to be an appropriate way to gain knowledge about an application domain for GIS purposes. The idea of conceptual learning is to construct classes of objects, where the typical properties are identified and collected. A class may then be defined as a structure of the common properties or components of all the instances (PART_OF-relation). Objects in these classes may have relations to other objects in a hierarchical manner (IS_A-relation). This shows a similarity with the entity-relationship approach and with structural and behavioural object-oriented database design, which makes it a usefull tool for the data modelling part in a GIS. Besides the procedural – or behavioural – approach in the application domain may as well be analysed and the knowledge about this may as well be captured using conceptual learning.

One may distinguish two types of knowledge acquisition approaches, which are the model-based and the incremental knowledge acquisition. In *model-based knowledge acquisition* the knowledge acquisition step is separated from the design and implementation phase. So the analysis of the expertise is independend from the knowledge representation later chosen for implementation. The result is a model of the expertise, its methods and their cooperation. In *incremental knowledge acquisition* the process of knowledge acquisition is regarded as an incremental, approximative and faulty process. The model of the expertise is produced during knowledge acquisition, which demands for a stepwise modelling, detection and correction of errors. An explorative prototyping – the prototype is only used for gaining the knowledge and designing a model of the expertise – is very usefull for that process of analysing the expertise making the design and implementation of the system independend from the prototype.

The design of the appropriate data model for the application and the design of the procedural flow including the adaptation of own methods on the GIS-product may be seen as two important steps where knowledge acquisition method could become very powerfull tools. In practise usually the step of knowledge acquisition and representation is very often ignored or executed in a less detailled way. Instead of modelling the application domain an immediate data structuring and procedural flow is implemented taking the specialised GIS-product characteristics into account. The disadvantage of this procedure is that the knowledge about the application domain is lost (or hidden). In particular one cannot easily recognize what part of the application domain is realized according to the specifications and what parts of the usual workflow is not supported because of system restrictions. The result of such an approach is very often that a solution exists, which is only partially related to the applications problem. It is proposed to use knowledge acquisition and representation methods for analysing and designing the application domain before implementing it on a GIS which is only able to manage procedural knowledge. The model-based knowledge acquisition is assuming a complete model of the application domain; it can be set up when the expert knowledge is transferred. Such an approach should be taken if the application domain is well known and production cycles are well established. The practical experience on the other hand shows that an incremental approach could become a better way to implement GIS in a new application field. But here usually the method of implementation is based on evolutionary prototyping, which means that at the end the prototype results in the production system. This very often has the disadvantage that production is restricted to the capabilities of the GIS.

The problem of all knowledge acquisition methods lies within the interpretation and analysis especially with respect to completeness and complexity. How does a human expert reach his problem solving feasibility ? Expertise also is based on human capabilities and implicit knowledge, that is not easy to query. Experts also rely on unconcious knowledge, even if they present it differently. Experts are experts in doing things not in explaining them.

2.4 Knowledge representation

The term 'knowledge representation' is defined by U. Reimer (1991) as the describing of symbols – so called representation structures – which fit in a recognizable manner to an extract of the world to be represented. Furtheron an interpretation instruction is needed which allows the formulation and evaluation of queries on that structure. A knowledge representation is standing for a set of circumstances – the represented part of the world – and is in itself a model of this part of the world. The knowledge base consists of the knowledge used by the system which is explicitely coded and is not implicitely hidden in the programming code. For further details in general see again U. Reimer (1991) and in a GIS-context see R. Bill (1991). There are two types of knowledge representation :

- Rule-based representation formats handle the rules in the form of statements. Well known representation schemata are models based on logic theory or on production rules. The whole knowledge is written down in the rules and statements; the data base itself is very often an unstructured and passive set of facts. In particular of interest for GIS is the fuzzy-logic which in comparison to the binary or boolean logic, where only true (0) and false (1) are allowed, supports a set of states. The fulfillment of a rule or statement may range in the intervall $[0, 1]$. An example of fuzzy-modelling is given in the application chapter in this paper (Figure 5). A further example for production rules is given there, too.
- Object-centered representation formats are placing the object itself in the foreground. Here ideas such as inheritance, methods etc. are included, these ideas are also found in object-oriented data modelling. Well-known types of object-centered representation formats are semantical networks, frames and scripts. For more detail see again U. Reimer (1991) in general and R. Bill (1991) with respect to GIS-applications. Semantical networks are ideal representations

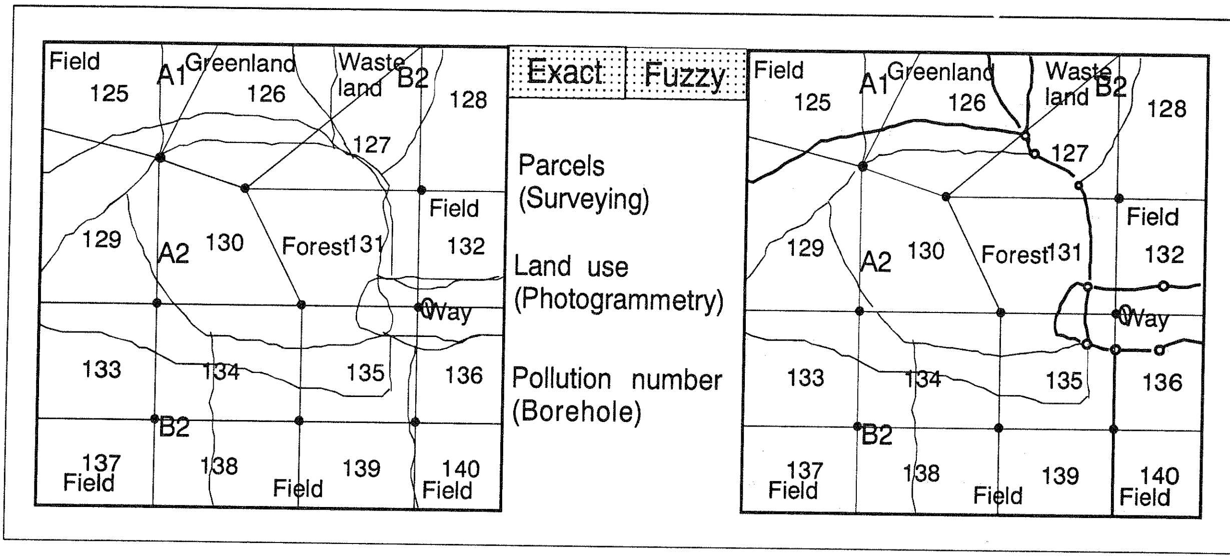

Figure 5 : Exact versus fuzzy polygon overlay.

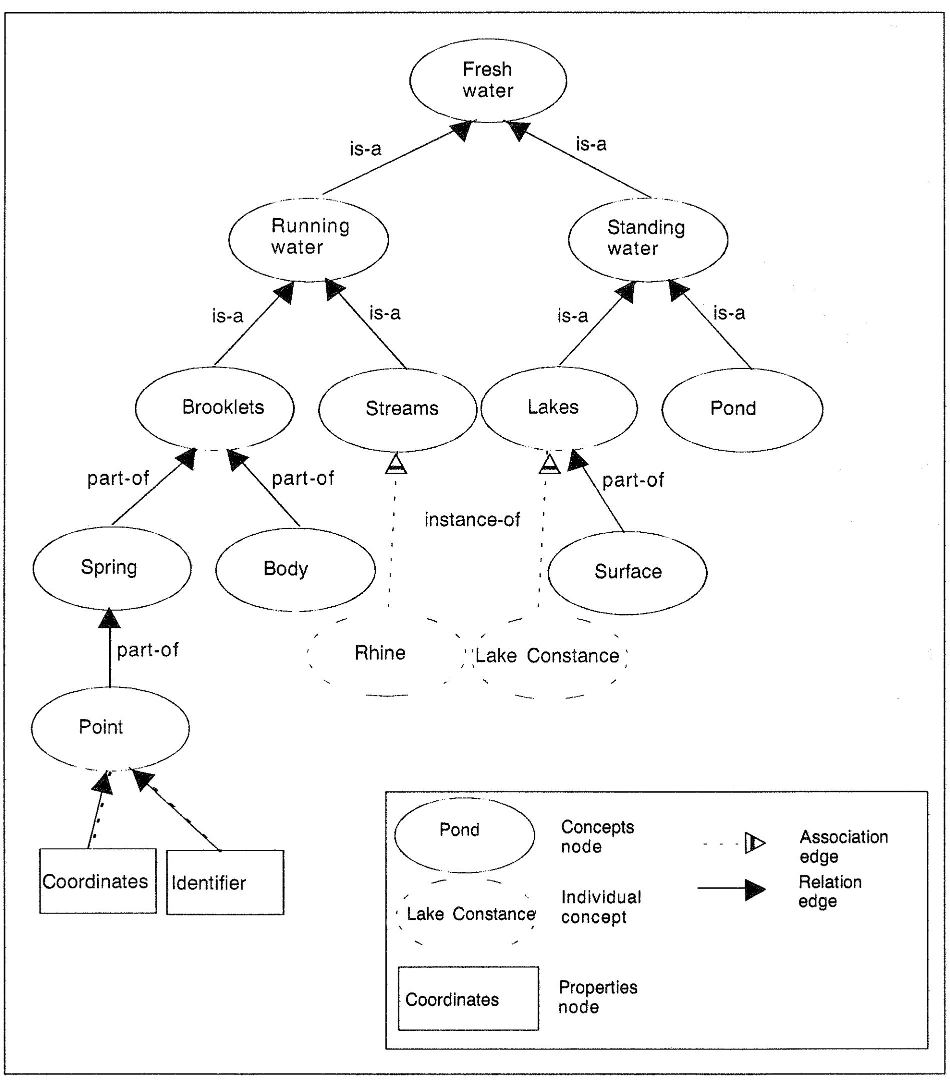

Figure 6 : A semantical network for fresh water.

for the GIS data modelling part. The entity-relationsship approach may be seen as a semantic network formalism. Analogies to the object-oriented method are given as well. An example for a data model of a GIS application based on the semantical network approach is given in the following chapter (Figure 6). Frames on the other hand are collecting the various properties of objects in slots. The individual values in a slot may be default values, methods, linkages to other frames or user given values. Frames are similar to complex objects currently investigated in database research as part of the structural object-orientation.

3 APPLICATION OF KNOWLEDGE IN GIS

Different GIS-applications where knowledge integration may be very effective should be illustrated now. G. Zhang and J. Tulip (1990) are presenting a fuzzy polygon overlay approach solving the sliver polygon problem quite reasonable. Figure 5 gives a comparison of an exact versus a fuzzy polygon overlay of three layers of information. The production rules may look like the following simplified statements :

- If many points resulting from intersection are identical within a certain tolerance they are merged together to one point.
- If many edges between points are identical within a certain bandwidth, these edges are merged and form one new edge.

Rule-based approaches may be suitable for raster-vector conversion and object recognition because of their capability to define rules with weak or fuzzy conditions. Usually scanned maps or images are no longer showing exact geometric and topological features, which causes severe problems for procedural and hard-coded approaches for conversion and object recognition. Further applications of rule-based methods may be given in rule-based assembly of features during data capture, for the production of thematic displays and for generalization purposes in GIS.

A last example illustrates the modelling of the real world structure and behaviour based on semantic networks. An extract of the real world – a model of fresh water features – is presented in figure 6. Semantical networks, frames and scripts are very well suited to GIS-data modelling.

4 REFERENCES

Bill, R., 1991. Konzepte der Wissens- und Datenrepräsentation in Geo-Informationssystemen. Proceedings of the 43 rd Photogrammetric Week at Stuttgart University. Schriftenreihe Heft 15. Institut für Photogrammetrie der Universität Stuttgart. pp. 99-113.

Bill, R., Fritsch, D., 1991. Grundlagen der Geo-Informationssysteme. Band 1 : Hardware, Software und Daten. Wichmann. Karlsruhe. 429 pages.

Bill, R., Fritsch, D., 1992. Grundlagen der Geo-Informationssysteme. Band 2 : Analysen, Anwendungen und neue Entwicklungen. Wichmann. Karlsruhe. ca. 429 pages.

Burrough, P.A., 1986. Principles of Geographic Information Systems for Land Resources Assessment. Oxford Science Publications. Monographs on Soil and Resources Survey No. 12. 194 pages.

Reimer, U., 1991. Einführung in die Wissensrepräsentation. B.G. Teubner. Stuttgart. 313 pages.

Rich, E., 1988. KI-Einführung und Anwendungen. Mc Graw Hill. Hamburg. 467 pages.