

KONZEPTE DER WISSENS- UND DATENREPRÄSENTATION IN GEO-INFORMATIONSSYSTEMEN

Ralf Bill, Stuttgart

1 Einführung

Der folgende Beitrag widmet sich dem Problem der Repräsentation von Daten und Wissen in einem Geo-Informationssystem. Bill/Fritsch /1/ definieren ein ' Geo-Informationssystem als ein rechnergestütztes System, das aus Hardware, Software, Daten und den Anwendungen besteht. Mit ihm können raumbezogene Daten digital erfaßt und redigiert, gespeichert und reorganisiert, modelliert und analysiert sowie alphanumerisch und graphisch präsentiert werden. ' Geo-Informationssysteme nutzen die Datenbanktechnik und kombinieren diese mit Informationssystemen. Sie sind prozedural ausgerichtet, d.h. Abläufe werden in Abhängigkeit von der Eingabe und den Daten von einem festcodierten Programm bewältigt. Damit stehen sie im Gegensatz zu den wissensbasierten Systemen, bei denen Daten, Wissen und Ablaufkontrolle eher deklarativ angelegt sind.

Im Beitrag werden vorweg einige Grundbegriffe diskutiert, wobei der Schwerpunkt auf der Wissenskomponente liegt. Die Daten- und Informationskomponente ist an vielen Stellen ausgiebig erläutert, so daß sie hier kürzer gefaßt werden kann. Die Bedeutung der Begriffe wird anhand von Beispielen aus dem GIS-Bereich illustriert. Im zweiten Teil wird der Versuch unternommen, die Vorteile der Wissensmodellierung und Wissensrepräsentation aufzuzeigen und mögliche Anwendungsbereiche im GIS zu skizzieren. Die gängigen Datenmodelle als Pendant auf der Datenseite werden ebenfalls kurz beschrieben. Sie haben bereits seit langer Zeit Einzug in die Geo-Informationssysteme gehalten.

1.1 Einige Grundbegriffe

1.1.1 Daten, Informationen und Wissen

Daten im klassischen Informatiksinne sind einfach nur Zeichen, die durch einen Computer gespeichert, verglichen, verarbeitet und geschrieben werden können. Erweitern wir diese Definition etwas, so können wir unter Daten auch Bilder, Texte, Graphiken, Sprache etc. verstehen. Daten selbst sind vom Computer interpretierbar, dem Menschen sagen sie ohne eine Interpretationsregel und eine Strukturierungsvorschrift relativ wenig.

Gängige Datenarten im Geo-Informationssystem (GIS) sind in Abbildung 1 angedeutet. Zu einem Objekt - einer für den Anwender des GIS bedeutungsvollen Abstraktion seiner Sicht der realen Welt - gehören seine Geometrie in Form von Punkten, Linien, Flächen oder Rasterzellen. Das Objekt besitzt weiterhin einen topologischen Typ und steht über topologische Beziehungen in Relation zu anderen Objekten. Neben dieser Form- und Beziehungsbeschreibung wird ein Objekt durch eine Vielzahl von Sachdaten (Attribute als Synonym) charakterisiert und besitzt unter einem bestimmten Thema eine definierte graphische Darstellung.

Der Begriff *Information* ist eng verknüpft mit Daten. Als Information wird zweckbezogenes Wissen bezeichnet. Hier soll darunter das Ergebnis der Anwendung von Transformationen, Regeln und Wissen von demjenigen verstanden werden, der mit den Daten vertraut ist, um daraus neue Fakten und interpretierbare Ergebnisse in einem gegebenen Rahmen zu erstellen. Die Kenntnis der Datenorganisation und des Datenzweckes gestattet es dem Nutzer, mit den Daten im Sinne von wertvollen und interpretierbaren Informationen zu arbeiten. Information kann in drei Ebenen aufgeteilt werden : die Syntax, d.h. die Zeichen, die verwendet werden, die Semantik, d.h. die Bedeutung und die Kommunikation, d.h. die Art und Weise wie Informationen verteilt werden. Insofern stellen die Objekte in einem GIS eigentlich schon Informationen - und nicht Daten - dar. Sie sind in der dem Benutzer spezifischen Syntax - seiner Sicht der realen Welt - niedergelegt und haben somit für ihn während der Bearbeitung direkt die gewünschte Bedeutung.

Objektart : Nadelwald

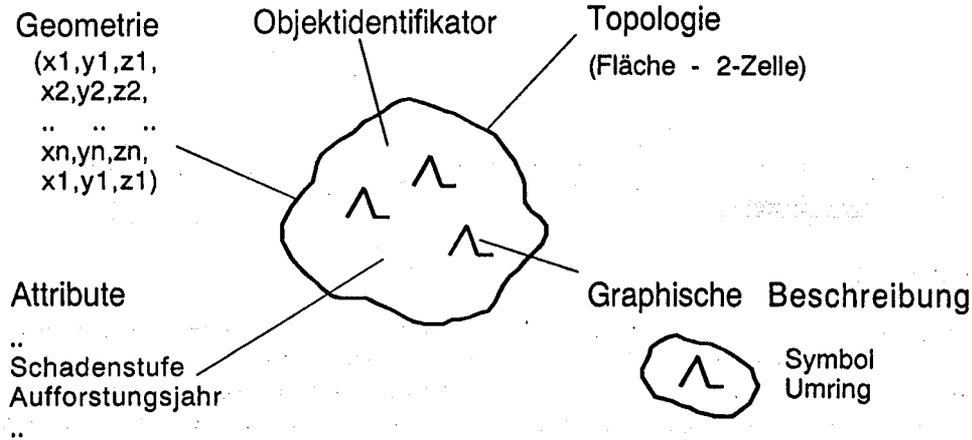


Abbildung 1 : Datenarten zu einem Objekt im GIS.

Als das *Wissen* eines Wissensträgers definiert Reimer /2/ die Menge aller von ihm als wahr angenommenen Aussagen über die repräsentierte Welt, die tatsächlich wahr sind. Demgegenüber stellen *Überzeugungen* eines Wissensträgers alle Aussagen dar, von denen er glaubt, sie seien wahr.

1.1.2 Wissensformen

Abbildung 2 zeigt verschiedene Formen von Wissen, die nicht als konkurrierend bzw. disjunkt aufzufassen sind sondern eher als sich gegenseitig ergänzende Formen.

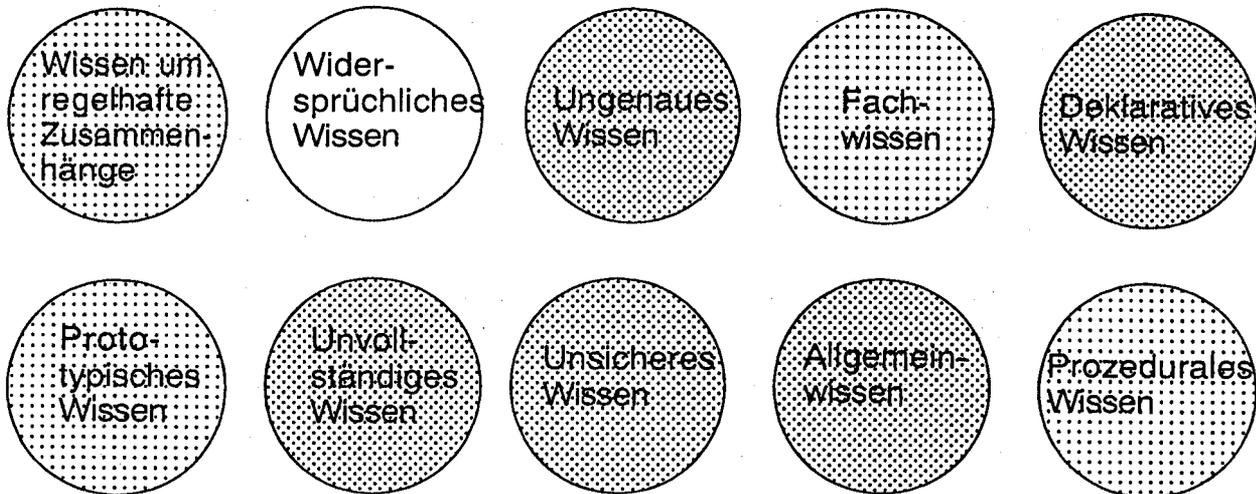


Abbildung 2 : Wissensformen.

Wir werden nur auf einzelne Formen näher eingehen und lehnen uns dabei erneut an die Darstellungen in /2/ an. Dem *prozeduralen Wissen*, d.h. Beschreibungen einer Vorgehensweise, wie man zur Lösung eines Problems gelangt, steht das *deklarative Wissen* gegenüber. Dieses beschreibt das Problem oder einen Sachverhalt, ohne auf die konkrete Vorgehensweise zur Lösung einzugehen. Zu deklarativem Wissen gehört auch immer ein Verfahren, das die Auswertung der Beschreibung übernimmt (Bruns /2/). So wird z.B. in einem 2D-GIS zu einem Gebäude dessen topologischer Typ als Fläche definiert, anhand dessen prozedural vom GIS gewährleistet wird, das jedes Gebäude im GIS auch als geschlossenes Polygon definiert ist.

Anhand eines Beispiels aus der Objektmodellierung in einem Geo-Informationssystem sollen die verschiedenen weiteren Wissensformen in Form von Aussagen spezifiziert werden (Abbildung 3). Das dargestellte semantische Netzwerk modelliert einen Ausschnitt aus einer baulichen Sicht auf den Datenbestand eines GIS. Die Objektklasse (im Sinne der Wissensrepräsentation als Konzeptklasse bezeichnet) 'Gebäude' mit den Unterklassen 'Industriegebäude' und 'Wohngebäude' (ebenfalls Konzeptklassen in einer Teil-von-Beziehung zur übergeordneten Objektklasse) seien betrachtet. Gebäude sollen durch einen Satz von Attributen (Merkmalen bzw. Eigenschaften in der Wissensrepräsentation) beschrieben werden. Diese Eigenschaften können an darunterliegende Konzeptklassen vererbt werden. Ein individuelles Objekt Gebäude wird datenbanktechnisch als Instanz und in Wissensrepräsentationen als Individuumkonzept bezeichnet, welches mit der es prägenden Konzeptklasse in einer Instanz-von-Beziehung steht.

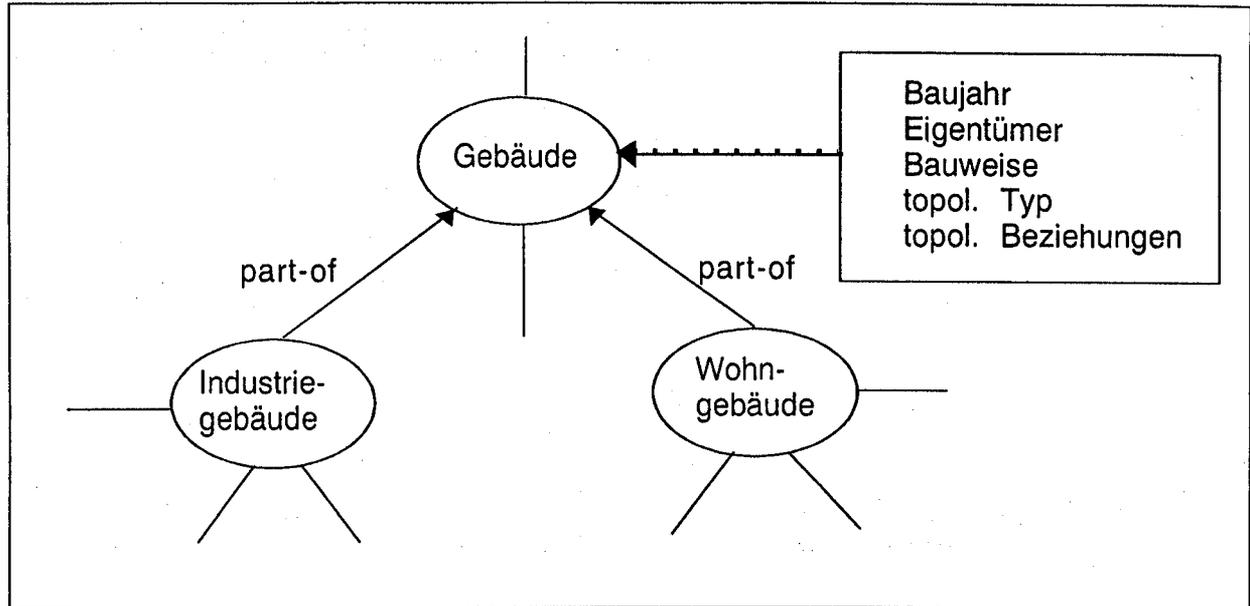


Abbildung 3 : Objektmodellierung im GIS.

Als *prototypisches Wissen* bezeichnen wir Wissen über normalerweise zutreffende Sachverhalte. Ausnahmen sind jedoch zugelassen und müssen durch weiteres Wissen spezifiziert werden. Aussagen der Form

- Gebäude besitzen ein Baujahr
- Gebäude haben einen Eigentümer
- Gebäude wurden in bestimmter Bauweise erstellt
- Gebäude sind disjunkt, d.h. sie überdecken sich nicht
- usw.

sind typische Aussagen für prototypisches Wissen. Diese Aussagen definieren das übliche Wissen über Gebäude; allerdings läßt sich für jede der genannten Aussagen auch eine Ausnahme finden (längere Bauperiode, mehrere Eigentümer oder Stockwerkseigentum, mehrere Bauweisen, verschachtelte Hochhäuser usw.). Die Modellierung der realen Welt in einem GIS beruht in der Regel auf solchem prototypischem Wissen, wobei als Ausnahmen allerdings nur exakt spezifizierte Fälle zugelassen sind, die das prozedurale System abarbeiten kann.

Widersprüchliches Wissen entsteht, wenn sich Aussagen selbst oder Aussagen den Fakten widersprechen. Ein Beispiel hierzu ist die Aussage

- Gebäude besitzt Baujahr vor 1900 und ist in Betonbauweise hergestellt.

Widersprüchliche Aussagen können in einem GIS generell nicht verarbeitet werden; der Nutzer muß den Konflikt bereinigen, bevor das System weiter agieren kann.

Unvollständiges Wissen zu dem Gebäude könnte die Aussage darstellen

- Gebäude ist Industriegebäude oder Wohngebäude.

Unsicheres Wissen stellt die folgende Aussage dar

- Gebäude ist möglicherweise ein Industriegebäude.

Ungenaueres oder unscharfes Wissen formuliert die Aussage

- Gebäude ist alt.

Objekte, auf die unvollständige, unsichere oder ungenaue Aussagen zutreffen, sollten in einem GIS grundsätzlich bearbeitbar und speicherbar sein. In der bis heute prozedural orientierten GIS-Welt ist dies allerdings nicht möglich.

Dagegen sind *regelhafte Zusammenhänge* als Wissen auch schon in den prozeduralen GIS verarbeitbar. Regelhafte Zusammenhänge bezeichnen Wissen der Art 'Wenn Faktum A zutrifft, dann gilt auch Faktum B'. Dafür können wir folgende Beispielaussagen formulieren.

- Wenn ein Objekt ein Industriegebäude ist, dann ist es auch ein Gebäude.
- Wenn ein Gebäude in Betonbauweise erstellt wurde, dann ist dessen Baujahr nach 1930.

Diese 'Wenn-Dann-Regeln' werden wir als klassisches Repräsentationsformat einer Produktionsregel in einem späteren Kapitel wiederfinden. Schlußfolgerungen aufgrund dieser regelhaften Zusammenhänge können auch heute schon durch implizite Codierung des Datenmodells und der Verarbeitungsalgorithmen in einem GIS gezogen werden.

Unter *Allgemeinwissen* versteht man Wissen, welches wir als Menschen einsetzen, um alltägliche Probleme zu lösen. Dieses Wissen läßt sich von *Fachwissen* abgrenzen, das Experten einsetzen, um Probleme spezialisierter Art zu lösen. Ein GIS basiert sehr stark auf Fachwissen und fachspezifischen Annahmen, welche wiederum implizit codiert sind.

1.2 Systeme

1.2.1 Datenbank- Managementsysteme

Datenbanken (im folgenden mit DB abgekürzt) kommt in Geo-Informationssystemen eine zentrale Bedeutung zu. Denn es ist der Datenbestand und die Flexibilität der Datenverwaltung, die bereits nach kurzer Zeit für den Nutzer von raumbezogenen Informationen den größeren Wert (auch in finanzieller Hinsicht) darstellen als die schnelle Hard- und Software.

Gegenüber der Dateienverwaltung der Vergangenheit, bei der Datenstrukturierung und Applikationsprogramm sehr eng miteinander verflochten waren, findet man - nicht nur im GIS-Bereich - zusehends Datenbanken als Speichermedien und Datenbank-Managementsysteme (DBMS) zur Manipulation der Daten. Aus Anwendersicht sollte eine Datenbank fähig sein, *Entitäten (Entity)* (Objekte) und deren Beziehungen (*Relationen (Relationship)*) in einem allgemeinen Modell zu repräsentieren und eine dem Benutzer angemessene Sicht auf diese Daten bieten. Diese Entitäten-Relationen-Modellierung muß flexibel sein und möglichst wenig Einschränkungen unterliegen.

Eine Entität ist ein bedeutungsvolles Primitiv - im Datenbanksinn die kleinste informationstragende Einheit - der darzustellenden Miniwelt; es ist ein Oberbegriff für Subjekt, Objekt, Ereignis oder Begriff. Die einzelnen Entitäten werden in Entitätenmengen zusammengefasst, welche gleichartige Eigenschaften (auch Attribute genannt) besitzen. Eine Entität ist innerhalb der Entitätenmenge durch einen Schlüssel - ebenfalls eine Eigenschaft - eindeutig charakterisiert. Beziehungen drücken einen bedeutungsvollen Zusammenhang zwischen den Entitäten aus. Während in der Datenbankwelt unter den Begriffen Entität und Beziehung wirklich diese kleinsten Informationseinheiten und deren Zusammenhänge zu verstehen sind, wird in der Anwendung GIS unter Entität (oder Objekt) schon ein in der realen Welt bedeutungsvolles Objekt (z.B. ein Haus) und unter dessen Beziehung z.B. der Zusammenhang zur Parzelle, auf der dieses Haus steht, verstanden. Als Eigenschaften (oder Attribute resp. Sachdaten) könnte die Entität Haus z.B. die Hausnummer, den Hausbesitzer, die Bewohnerzahl etc. tragen. Diese notwendige Atomisierung auf der Einfügeseite (Insert) - Zerlegung der Realweltobjekte in Datenbankatome - und die Aggregation auf der Rückgewinnungsseite (Retrieval), um von den Datenbankatomen zu in der realen Welt bedeutungsvollen Objekten zu kommen, führt allerdings zu Problemen der gängigen Datenbanken mit GIS-Anwendungen (vgl. Bill /4/).

1.2.2 Informationssysteme

In seiner einfachsten Form ist ein Informationssystem ein Frage-Antwort-System auf einen Datenbestand. Ein Informationssystem dient dazu, Informationen für Zwecke der unterschiedlichsten Art vorrätig zu halten und bei Bedarf zur Verfügung zu stellen. Es erlaubt durch spezifische Verknüpfungen unter den Daten die Gewinnung neuer Informationen. Informationssysteme gehören heute in vielen Bereichen zu Standardwerkzeugen (Management-Informationssysteme, Flugreservierungssysteme etc.). Ein Informationssystem ruht auf den vier Säulen Hardware, Software, Daten und den Anwendern, welches zum sogenannten Vierkomponenten-Modell HSDA führt (/1/). Ein Geo-Informationssystem ist als ein spezielles Informationssystem zu betrachten, welches darüberhinaus noch die vier Komponenten EVAP - nämlich die Erfassung, die Verwaltung, die Analyse und die Präsentation raumbezogener Daten - unterstützen muß.

1.2.3 Wissensbasierte Systeme

Die Hauptaufgaben wissensbasierter Systeme liegen im Erkennen von Beziehungen, dem Ziehen von Schlußfolgerungen und dem Steuern der Anwendung von Wissen (Kleine Büning /5/). Sie benötigen hierzu Techniken der Wissensrepräsentation, der Wissensverarbeitung und des Wissenserwerbs. Wir gehen hier nur auf die Methoden der Wissensrepräsentation ein. Expertensysteme sind der derzeit expandierende Zweig der wissensbasierten Systeme.

Die Wissensbasis stellt die Menge der vom wissensbasierten System verwendeten 'Wenn-Dann-Aussagen' zusammen, mit deren Hilfe automatisierte Folgerungen abgeleitet werden können. Formal entspricht sie der Datenbasis in einem Datenbanksystem. Die Wissensverarbeitung erfolgt allerdings vollständig anders. Ein Inferenzmechanismus, d.h. ein strukturiertes Verfahren, mit dem das Wissen einer Repräsentationsform analysiert und verknüpft werden kann, ermöglicht das Ziehen von Schlußfolgerungen aufgrund gegebener Zusammenhänge.

2 Repräsentationsschemata für Daten

2.1 Beziehungarten

Zwischen den Entitäten bzw. Objekten können die folgenden Beziehungen auftreten :

- die 1:1 Beziehung : Einem Objekt des Typs A ist umkehrbar eindeutig ein Objekt des Typs B zugeordnet.
- die 1:n (oder n:1) Beziehung : Einem Objekt des Typs A sind beliebig viele Objekte des Typs B zugeordnet (oder umgekehrt).
- die m:n Beziehung : Es besteht eine Zuordnung zwischen beliebig vielen Objekten des Typs A und Objekten des Typs B.

2.2 Logische Datenmodelle und ihre Anwendung im GIS

Das logische Datenmodell beinhaltet die Festlegung von Datenstrukturen zur Verwaltung der Daten - oder der speziellen Sicht auf die Daten - und die zugehörigen erlaubten Operatoren auf diese Strukturen. Das logische Datenmodell ist die Basisinformation eines GIS, nicht die physikalische Datenspeicherung. Die Realwelt beschreiben wir im logischen Modell mit Objekten, den Beziehungen zwischen der Realwelt und den modellierten Objekten und den Beschreibungsregeln für beide. Im folgenden Abschnitt sollen die wesentlichen Arten der logischen Datenmodellierung - wir unterscheiden hier zwischen hierarchisch, netzwerkartig, relational und objektorientiert - diskutiert werden.

2.2.1 Das hierarchische Datenmodell

Hierarchische Datenmodelle sind die ältesten der logischen Modelle; sie setzen eine Hierarchie in den Beziehungen der Daten voraus und nutzen diese bei der Verwaltung aus. So sind z.B. baumartig organisierte Datenstrukturen (binär, quadtree ..) hierarchische Datenmodelle. Die Vater-Sohn Beziehungen werden von den Operatoren explizit genutzt. Operationen navigieren durch die Baumstruktur. Das hierarchische Datenmodell unterstützt relativ einfach die 1:n Beziehung, während m:n Beziehungen nur durch redundante Speicherung von Information modelliert werden kann. Jedes hierarchische Datenmodell erlaubt eine sehr rationelle Einreihung der Daten (sequentiell) und ist sehr effizient, wenn den Daten diese Eigenschaft zugrundeliegt.

Man spricht beim hierarchischen Modell von :

- ROOT - Stamm, Ausgangskriterium
- LEAVES - Blätter, Folgemerkmale

Hierarchische Modelle sind kompliziert zu handhaben, da Relationen durch eine beschränkte Anzahl von Links (Verknüpfungen zwischen Records) dargestellt sind. Man könnte sich das Verwalten der Daten in einfachen Listen oder sequentiellen Dateien vorstellen. Verfeinerungen sind geordnete Listen und indexsequentielle Dateien.

Ein Beispiel (nach Burrough /6/, vgl. auch Güntsch /7/ und Bill /4/) zeigt die Verwaltung von zwei Polygonen in einem hierarchischen Modell (Abbildung 4).

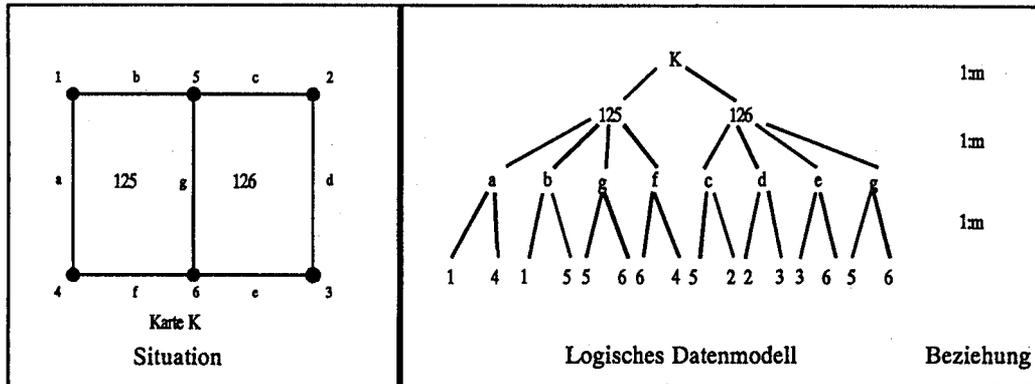


Abbildung 4 : Beispiel eines hierarchischen Datenmodells im GIS-Bereich.

Die Verwaltung von Rasterdaten beruht oftmals auf einer hierarchischen Zellteilung in Form des Quadtree.

2.2.2 Das netzwerkartige Datenmodell

Der Begriff netzwerkartig reflektiert im wesentlichen den Beziehungsteil der Realwelt. Man kann es als verallgemeinerte Methode des hierarchischen Modells betrachten mit einer beliebigen Anzahl von Verknüpfungen. Das netzwerkartige Datenmodell ist durch den Data Base Task Group-Vorschlag der CODASYL /8/ standardisiert. Eine solche Datenbank wird auch CODASYL-Datenbank genannt.

Beim netzwerkartigen Modell spricht man direkt von

- RECORDS - den Objekten
- LINKS oder SETS - den Beziehungen

Es ist charakterisiert durch Owner-Member-Beziehungen. Ein Owner-Record-Typ ist ein Objekt, dem andere Datenobjekte - Member-Record-Typen - über einen Set-Typ in fester Reihenfolge nachgeordnet sein können. Einem Owner können viele Member derselben Art, ein Member kann mehreren Ownern über jeweils ein Set zugeordnet sein. Jeder Set-Typ besitzt einen Owner- und einen Membertyp, womit indirekt auch m:n Beziehungen realisiert werden. Rekursive Strukturen sind schwer zu behandeln. Operationen in netzwerkartigen Modellen sind charakterisiert durch das Navigieren auf Set-Typen. Aus einem netzwerkartigen Datenmodell kann durch redundante Daten ein hierarchisches Modell erzeugt werden. Problematisch ist bei netzwerkartigen Modellen die Erweiterung des Modells.

Netzwerkartige Datenmodelle findet man häufig in der Geometrie- und Graphikdatenverwaltung von GIS, z.B. auch bei SICAD für die Verwaltung von Punkten, Linien, Flächen, Texten und benutzerdefinierbaren Elementen.

Auch hier sei wieder das Beispiel der Polygonverwaltung aufgezeigt (Abbildung 5).

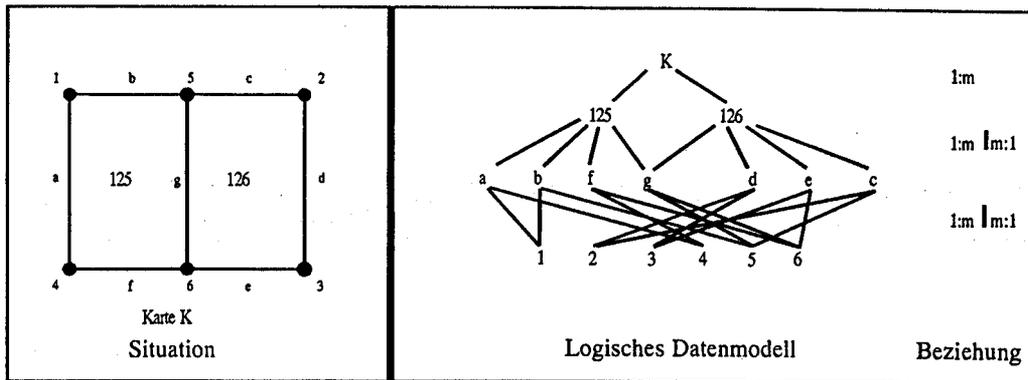


Abbildung 5 : Beispiel eines netzwerkartigen Datenmodells im GIS-Bereich.

2.2.3 Das relationale Datenmodell

Das relationale Modell überträgt die Ergebnisse der mathematischen Theorie der Relationen auf die Datenbankproblematik. Es ist im Konzept auf Tabellen aufgebaut, in denen die Daten und Beziehungen zwischen den Daten verwaltet werden. Im Umfeld des relationalen Modells arbeitet man mit den Begriffen

- **RELATION** = Menge von Tupeln (Daten und Beziehungen zwischen Daten, deren Umsetzung in eine Tabelle mit Tabellename gleich Relation), die durch Angabe des Namens und der beteiligten Attribute festgelegt ist,
- **TUPEL** = geordneter Satz von Attributen, untereinander in zweidimensionalen Tabellen eingetragen, in Tabellensicht eine Zeile einer Tabelle mit einem eindeutigen Schlüssel auf das Tupel,
- **ATTRIBUT** = ein Eintrag zu einem bestimmten Merkmal (Entität oder Beziehung) einer Relation, also in Tabellensicht ein Spaltenwert,
- **DOMÄNE** = Wertebereich oder die Menge der verschiedenen Feldwerte eines Attributes,
- **SCHLÜSSEL** = ein Attribut (oder eine Kombination von Attributen) mit der Eigenschaft, daß in allen möglichen Ausprägungen der Relation jeder Wert nur einmal vorkommt. Man unterscheidet Primär- und Fremdschlüssel.

Aus der Sicht relationaler Datenmodelle bearbeitet man eine Datenbank mit den relationalen Operatoren :

- **Projektion** = Auswahl von Attributen aus einer Tabelle.
- **Selektion** = Auswahl von Tupeln, die einer bestimmten Bedingung genügen.
- **Verknüpfung** = das kartesische Produkt von Zeilenmengen.
- **Produkt** = das Produkt einer Spalte mit einer anderen.
- **Vereinigung** = die mengenmäßige Vereinigung zweier Teilmengen.
- **Differenz** = der Unterschied zwischen zwei Teilmengen.
- **Durchschnitt** = die Gemeinsamkeiten zweier Teilmengen.
- **Dividieren** = das Aufsplitten eines Produktes.

Dieses auf Codd /9/ zurückzuführende Modell verdankt seine Verbreitung der klaren mathematischen Darstellung und dem am strengsten realisierten Entkoppeln der Datenbereitstellung von der Anwendung, d.h. die strikte Trennung von der physikalischen Speicherung und der logischen Modellierung. Die Tabelle ist die einzig gültige Datenstruktur. Den Objekten und Beziehungen werden Attribute zugeordnet, die als atomare (unstrukturierte) Einheiten betrachtet werden.

Bemerkenswert am relationalen Modell ist, daß Beziehungen zwischen Objekten ebenfalls als Werte in Tabellen, also wie die Objekte selbst verwaltet werden. Entitäten und Relationen werden in einer Weise repräsentiert. Außerdem ist die Tabellenform leicht zu verstehen und mit einem einfachen Operationsset bedienbar. Das relationale Modell ist allerdings schlecht für den räumlichen Zugriff geeignet. Unterhalb der logischen Modellierung im relationalen Modell kann die physikalische Umsetzung im Speicher der Rechner wiederum auf hierarchische oder netzwerkartige Strukturen aufbauen. Für den Entwurf eines relationalen Modells und dessen Umsetzung sind Normalformen anzustreben. Nicht normalisierte Relationen können Attribute besitzen, deren Werte wieder Relationen sein können. Man spricht von 1., 2., 3. usw. Normalform, je nachdem wie weit die Normalisierung der Entitäten und Relationships fortgeschritten ist. Die 1. Normalform ist erhalten, wenn eine Relation nur noch einfache Attribute enthält; d.h. in jeder Tabellenspalte steht nur noch ein Wert, ein Atom. Höhere Normalformen sind anstrebenswert, um funktionale Unabhängigkeit der Attribute zu erreichen.

Kommerzielle Datenbanken aufbauend auf dem relationalen Modell sind heute in großer Vielzahl erhältlich. Ein Problem der relationalen Datenbanken ist ihr Antwortzeitverhalten bei komplexeren Anfragen. Aus diesem Grund finden sie auch bisher nur in sehr wenigen GIS Einsatz (Beispiel System 9 von Computervision GIS AG oder Gradis UX von STI) als Verwaltungssysteme für Geometriedaten; verstärkt sind sie allerdings vertreten bei der Verwaltung der beschreibenden Informationen, die eine hohe Ähnlichkeit mit denjenigen Daten besitzen, für die das relationale DBMS entworfen wurde (Beispiel MGE von Intergraph u.a.).

Auch hier sei wieder das Beispiel der Polygonverwaltung aufgezeigt (Abbildung 6).

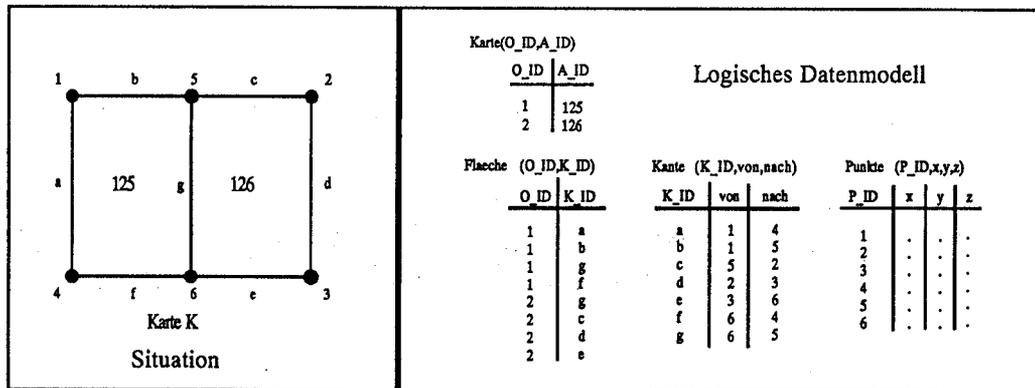


Abbildung 6 : Beispiel eines relationalen Datenmodells im GIS-Bereich.

2.2.4 Das objektorientierte Datenmodell

Derzeit sind in der Informatik Forschungen auf dem Gebiet der Erweiterung des relationalen Datenmodells zu objektorientierten Datenmodellen im Gange. Ziel ist hierbei die Unterstützung neuer Datentypen durch Aufgeben der ersten Normalform, die Unterstützung der dazugehörigen Operatoren und die Implementation raumbezogener Speicher- und Zugriffsmechanismen (Schek /10/).

Hinsichtlich der objektorientierten Datenmodelle kann man drei Klassen unterscheiden (Dittrich /11/):

- Strukturell objektorientierte Datenmodelle.
- Verhaltensmäßig objektorientierte Datenmodelle.
- Voll objektorientierte Datenmodelle.

Diese unterscheiden sich hinsichtlich der Flexibilität. Strukturell objektorientierte Datenmodelle können komplexe Objekte als ein Atom in der Datenbank behandeln, wobei die Operatoren, die auf diese Objekte angewendet werden können, vordefiniert sind. Verhaltensmäßig objektorientierte Datenmodelle erlauben anwendungsspezifische Datentypen und Operatoren, abstrakte Datentypen und Integritätsbedingungen. Voll objektorientierte Datenmodelle sind strukturell und verhaltensmäßig objektorientiert.

Zum Verständnis des objektorientierten Konzepts seien einige Begriffe erläutert /7/ (Abbildung 7). Ein Objekt besteht aus einem Datensatz und einer Anzahl von Operationen (methods), die es ausführen kann. Ein Objekt kann über Kanäle (channels) mit anderen Objekten verbunden sein. Diese Beziehungen können vom Typ 1:1, 1:n

oder m:n sein. Ein Objekt kann nur über einen seiner Kanäle mit einer Nachricht (message) angesprochen werden und zu einem Einsatz seiner Operationen veranlaßt werden. Der Benutzer sieht nur die Ausführung, nicht die interne Realisierung. Das Objekt ist eine kleine Welt für sich, weshalb man auch von der Einkapselung (encapsulation) spricht. Jedes Objekt gehört einer Objektklasse an, aus der es durch die Festlegung bestimmter Attribute entstanden ist. Erst bei Vorliegen aller Attribute wird es zu einem Individuum (instance). Durch Vererbung von Eigenschaften (inheritance) können neue Objektklassen gebildet werden. Die daraus gebildete Unterklasse erbt alle Funktionen und Daten dieser Oberklasse und es können Methoden umdefiniert und neue ergänzt werden. Die Vorteile des objektorientierten Ansatzes liegen klar auf der Hand. Das Objekt ist flexibel definierbar und änderbar als Abbild eines Bestandteils der realen Welt; es besitzt ein Eigenleben und verwaltet seine Daten selbständig. Der Nachteil ist allerdings neben der fehlenden bisherigen Verfügbarkeit solcher Datenbanken deren Antwortzeitverhalten und das rasche Anwachsen von Quellcode. Rechenaufwand und Speicherbedarf ist extrem. Dennoch wird dieses Modell in Zukunft bedeutend sein. Im GIS-Bereich gibt es noch wenige Produkte (TIGRIS von Intergraph und Smallworld GIS von Smallworld), die als weitestgehend objektorientiert bezeichnet werden können.

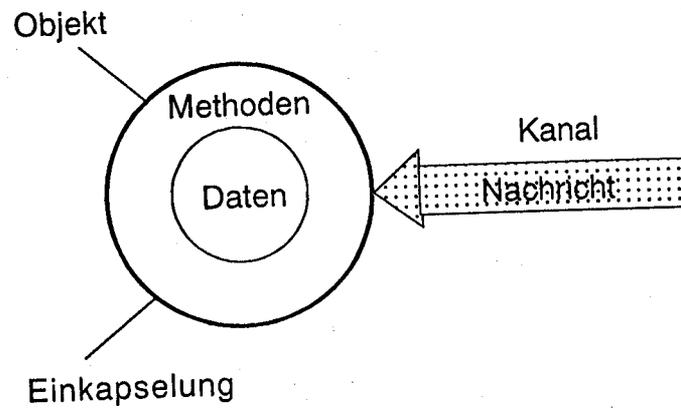


Abbildung 7 : Grundbegriffe des objektorientierten Ansatzes.

Auch hier sei wieder das Beispiel der Polygonverwaltung aufgezeigt (Abbildung 8).

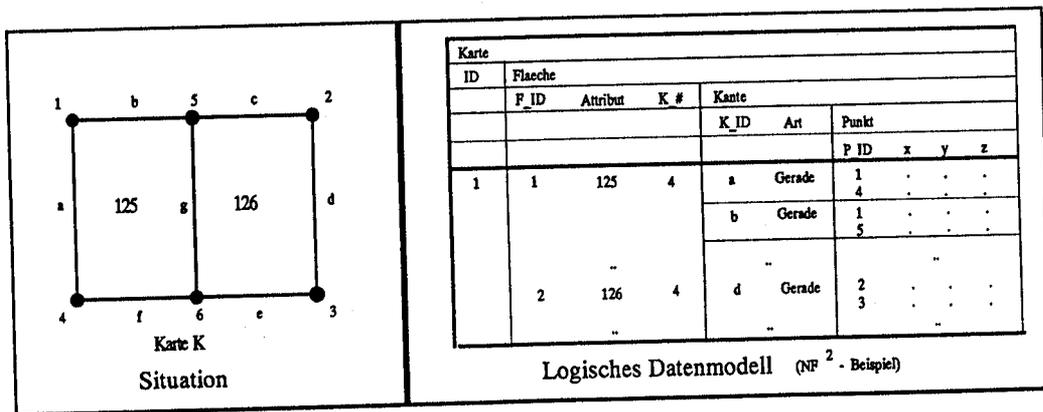


Abbildung 8 : Beispiel eines objektorientierten Datenmodells im GIS-Bereich.

2.2.5 Das HyperMedia Konzept

Als neue Entwicklung für den Bereich der Datenverwaltung ist das HyperText-Konzept anzusehen (Seidensticker /12/), welches auf dem Begriff des semantischen Netzes aufbaut. Semantische Netze sind frei definierte Graphen. Ein Graph besteht aus Knoten und Kanten, die Knoten untereinander verbinden. Die Knoten sind hierbei Träger von Objekten, die Kanten stehen für Beziehungen zwischen den Objekten. Bei HyperText sind die Objekte textuale Informationselemente, während in der Verallgemeinerung - dann als HyperMedia bezeichnet - auch Komponenten wie Graphik, Sprache, Bilder, Musik und Animation als Objekte integriert sind. Die Bezeichnung 'Hyper' leitet sich davon ab, daß dieser Graph zur Darstellung von Beziehungen gegebenen Datenbeständen logisch gesehen übergestülpt wird. Die Beziehungen werden sowohl innerhalb eines Dokuments (im Falle von

HyperText) als auch zwischen Dokumenten durch maschinen-unterstützte Verbindungen (links) realisiert. Somit lassen sich komplexe Datenbestände und deren Beziehungen separat erfassen. Zugriffsmöglichkeiten sind damit nicht nur über das Datenmaterial sondern auch über die zwischen ihnen festgelegten Beziehungen möglich. Erste Prototypen von Hypertext gibt es seit Ende der 60 er Jahre. HyperText-Umgebungen sind heute als Produkte verfügbar (HyperCard und GUIDE) für Apple Macintosh-Computer. Eine Verschmelzung einer Oracle-Datenbank mit einem HyperTalk-Programm auf dem Macintosh im Bereich GIS zeigt Gong /13/. Dieses Konzept wird in Zukunft besonders für Multimedia-GIS an Bedeutung gewinnen. Auf die Bedeutung semantischer Netze als Wissensrepräsentationsformate werden wir später eingehen.

3 Repräsentationsschemata für Wissen

Den Begriff *Wissensrepräsentation* definiert Reimer /2/ als das Aufschreiben von Symbolen (Repräsentationsstrukturen), die in einer erkennbaren Weise einen Ausschnitt einer zu repräsentierenden Welt entsprechen. Das Aufschreiben dieser Strukturen ergibt erst dann eine Repräsentation, wenn weiterhin eine Interpretationsvorschrift vorliegt, die mindestens die Formulierung und Auswertung von Anfragen auf diesen Strukturen zuläßt. Eine Wissensrepräsentation steht somit stellvertretend für eine Menge von Sachverhalten - die repräsentierte Welt - und ist damit ein Modell für diese Welt. Die Wissensbasis beinhaltet das vom System benutzte Wissen, welches explizit codiert ist und damit nicht implizit im Programmcode versteckt ist. Durch diese explizite Codierung wird das Wissen austauschbar, erweiter- und modifizierbar ohne Änderung des Programmcodes. Die Wissensbasis besteht aus Informationen über Objekte und Regeln. Ein Objekt beschreibt dabei ein Phänomen, das durch eine Menge anwendbarer Regeln und Attribute definiert ist. Die Regeln definieren die Operationen, die mit oder von einem Objekt durchgeführt werden können, sofern bestimmte Bedingungen erfüllt sind. Attribute spezifizieren die Eigenschaften eines Objektes näher. Je nachdem, ob die Wissensbasis stärker durch den Informationstyp Objekt oder Regel beschrieben wird, spricht man von objekt-zentrierten oder regelbasierten Systemen (Abbildung 9).

Die Datenmodellierung im GIS definiert ein Abbild der realen Welt. Das dazugehörige Bearbeitungsmodell ist allerdings im klassischen Sinne prozedural abgelegt. GIS verfügen damit über keine Wissensbasis, sondern besitzen eine direkte Codierung des problemspezifischen (domänenabhängigen) Wissens. Heutige GIS können am ehesten mit objekt-zentrierten Systemen bei fest definierten Regeln verglichen werden. Diese Einschränkung steht symptomatisch für die Restriktionen, die ein GIS dem Benutzer i.d.R. auferlegt. Das Geo-Informationssystem ist in dem Sinne ein traditionelles Computerprogramm, welches zwischen den Daten und den Algorithmen unterscheidet. Die einmal im Programmcode festgelegten Algorithmen definieren den Umgang mit den veränderlichen wachsenden Datenbeständen. Die Daten sind in einem starren Schema (z.B. Tabellenform mit gleichartiger Satzstruktur, fester Feldlänge und eindeutigem Schlüssel, vgl. logische Datenmodelle) eingezwängt.

3.1 Regelbasierte Wissensrepräsentationsformen

Regelbasierte oder nicht-objektzentrierte Repräsentationsformen stellen die Regeln in Form von Aussagen in den Vordergrund. Bekannteste Vertreter sind auf Logikmodellen oder Produktionsregeln aufgebaut. Das gesamte Wissen steckt in den Regeln. Die Datenbasis ist oftmals nur eine unstrukturierte und passive Menge von Fakten. Die Bestandteile eines solchen regelbasierten Systems sind

- die Datenbasis mit allen gültigen Fakten
- die Regeln zur Herleitung neuer Fakten und Aktionen
- die Regelinterpreter zur Steuerung des Herleitungsprozesses.

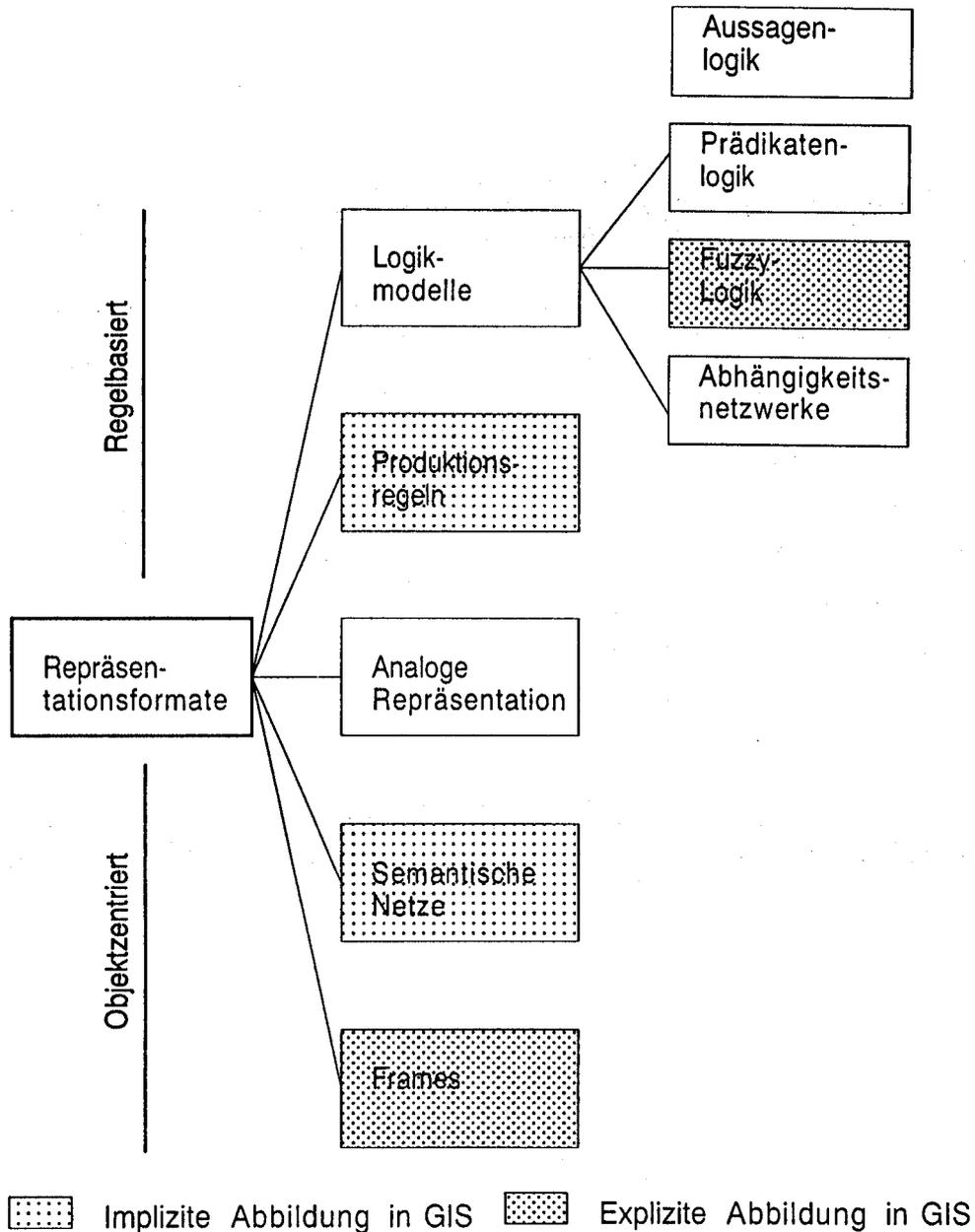


Abbildung 9 : Wissensrepräsentationsschemata.

Nicht-objektzentrierte Repräsentationsformen sollen wieder an einem Beispiel aus der GIS-Welt illustriert werden. Die *Aussagenlogik* benutzt deklarative Aussagen in der Form

- Ein Gebäude hat einen Eigentümer.
- Ein Wohnhaus ist ein Gebäude.

Logische Verknüpfungen durch UND, ODER, NEGATION und IMPLIKATION sind möglich. Die UND-Verknüpfung führt mit obigen Aussagen zu dem Ergebnis

- Ein Wohnhaus ist ein Gebäude und hat einen Eigentümer.

Die Aussage dieser Verknüpfung besitzt einen Wahrheitswert (wahr oder falsch). Allerdings sind keine Verallgemeinerungen möglich. Dies bleibt der *Prädikatenlogik* vorbehalten, in der statt einfacher Aussagen Prädikate verwendet werden. Die Prädikatenlogik beinhaltet Operatoren zur Existenzaussage (\exists) und zur Verallgemeinerung (\forall). Mit dem Funktionssymbol 'Inklusion' und den Konstanten 'Gebäude', 'Parzelle', 'Eigentümer' und 'Fläche' legen die folgenden Formeln die Semantik fest.

- Inklusion(Gebäude) = Parzelle
- hat-Teil (Gebäude, Eigentümer)
- hat-Teil (Parzelle, Eigentümer)
- $\exists x := (\text{Inklusion}(x), \text{Parzelle})$

Unschärfes Wissen kann mit den sehr scharfen (binären) Aussagen der Logikmodelle nur bedingt behandelt werden. Diese können mit mehrwertigen Logiken wie z.B. der Fuzzy-Logik repräsentiert werden. Statt der 0,1-Aussagen sind eine eingeschränkte Menge von Wahrheitswerten zugelassen, die mit Teilmengen aus dem Intervall [0,1] identifiziert werden. Erste Ansätze zur Nutzung von Fuzzy-Modellen in GIS sind in System 9 im Polygon Processing zu finden (Zhang/Tulip /14/). Das Ergebnis einer Flächenverschneidung von drei Informationsebenen sehr unterschiedlicher geometrischer Qualität - Flurstücke, Landnutzung und Schadstoffgehalt - stellt Abbildung 10 dar. Dabei ist der exakten Flächenverschneidung das Resultat einer Fuzzy-Flächenverschneidung gegenübergestellt. Mit derartigen Ansätzen kann das Problem der 'Sliver-Polygone' - das sind kleine, durch die Verschneidung entstehende Polygone ohne fachliche Bedeutung - vermindert werden. Dies ist deutlich im rechten Bild erkennbar, in dem die vielen Kleinformen aus dem linken Bild nicht mehr auftreten. Die exakte Geometrie wird durch Punktklassenbildung und Kantenverschmelzung aufgehoben.

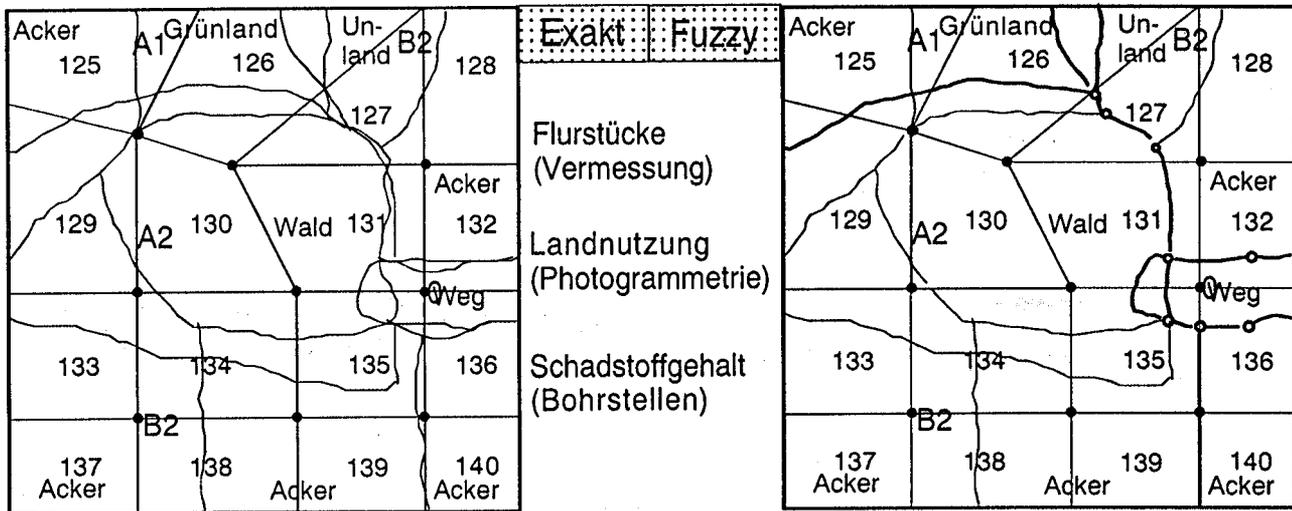


Abbildung 10 : Exakte versus Fuzzy-Flächenverschneidung.

Produktionsregeln eignen sich am besten zur Darstellung regelhafter Zusammenhänge. Eine Produktionsregel beschreibt eine mit einer Vorbedingung versehene Aktion. Diese Aktion gilt als ausführbar, wenn die Vorbedingung erfüllt ist. Die Vorbedingungen beziehen sich auf die Faktenbasis in beliebigem Repräsentationsformat. Die Fakten stellen die bis jetzt als gültig erachteten Tatsachen dar. Die Regeln beschreiben die Beziehungen zwischen den einzelnen Tatsachen in 'Wenn-Dann-Aussagen'. Der 'Wenn-Teil' entspricht der Vorbedingung; der 'Dann-Teil' beschreibt die auszuführende Aktion. Als Produktionsregeln für das vorgenannte Fuzzy-Polygon-Overlay könnten - vereinfachend - die beiden folgenden Regeln zur Anwendung kommen.

- Wenn mehrere Punkte aus der Verschneidung innerhalb einer vorgegebenen Toleranz identisch sind, dann verschmelze diese zu einem Punkt.
- Wenn mehrere Kanten zwischen Punkten innerhalb einer vorgegebenen Toleranz gleich verlaufen, dann verschmelze diese zu einer Kante.

Für Produktionsregeln ist ein großes Anwendungsspektrum im GIS zu sehen; zu nennen sind hier regelbasiertes Objekterzeugen (Ruled based feature assembly in System 9), die Objekterkennung aus gescannten und vektorisierten Karten, die Generalisierung u.a..

3.2 Objektzentrierte Wissensrepräsentationsformen

Bei objektzentrierten Repräsentationsformen steht das Objekt (und nicht die Regel) im Vordergrund. Zentrale Ideen solcher Ansätze sind die Vererbung, Defaultwerte und zugeordnete Methoden. Bekannte Formen

der Repräsentation sind semantische Netze und Frames /2/,/15/.

Semantische Netze sind markierte Graphen. Die Knoten stellen die Repräsentation von Objekten, Ereignissen, Situationen etc. dar. Die Kanten im Graph stellen die Beziehungen zwischen den Knoten dar. Semantische Netze sind Mittel zur graphischen Repräsentation von Wissen und weniger zur maschinellen Verarbeitung von Wissen geeignet. Abbildung 11 stellt einen Auszug eines GIS-Datenmodells als semantisches Netz dar. Insbesondere für den Schritt der Datenmodellierung sind semantische Netzwerke ideal auch im GIS-Bereich einsetzbar. Das Entitäten-Relationenmodell (vgl. Kapitel 1.2.1) kann als semantischer Netzformalismus angesehen werden. Dies ist im Datenbankbereich entstanden und findet Anwendung bei der Datenmodellierung. Erste Ansätze einer Anwendung semantischer Netze zur Datenmodellierung zeigen sich im System Tigris von Intergraph.

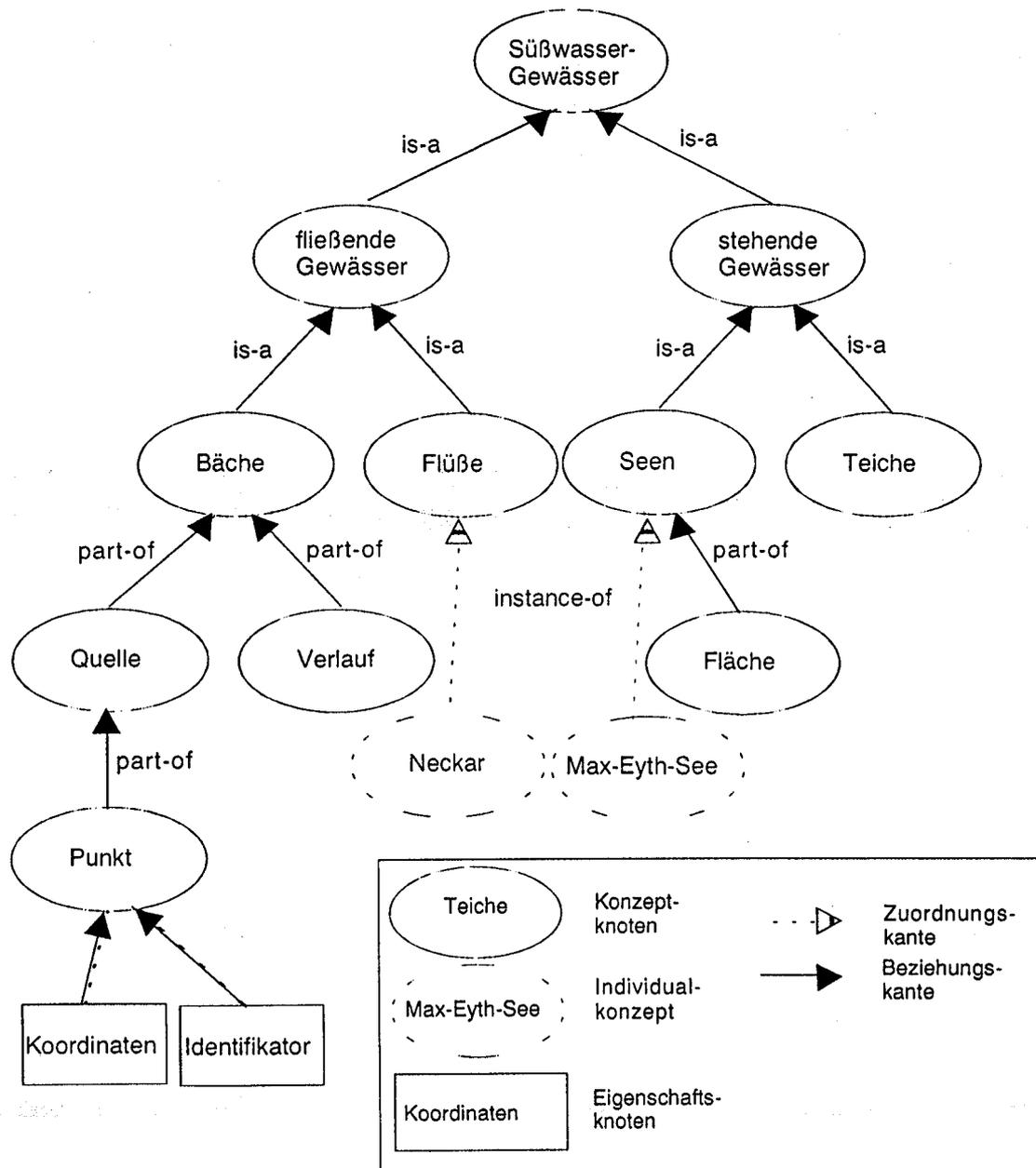


Abbildung 11 : Ausschnitt eines semantischen Netzwerkes für Gewässer.

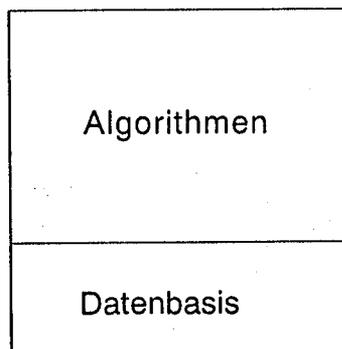
Frames fassen verschiedene Eigenschaften eines Objektes in einem Rahmen zusammen. Die Eigenschaften sind in Slots gespeichert, denen die Eigenschaftswerte zugewiesen werden. Die individuellen Slotwerte können Defaultwerte, Methoden, Verweise auf andere Frames oder Benutzerwerte sein. Mit Frames im weiteren Sinne vergleichbar

sind die im Datenbankbereich aufkommenden komplexen Objekte. Komplexe Objekte und Frames sind beides objektzentrierte Repräsentationsansätze. Diese auch als strukturelle Objektorientiertheit bezeichneten Konzepte können durch Einbeziehung von Methoden zu verhaltensmäßig orientierten Ansätzen erweitert werden.

4 Schluß

Die Kenntnisse und Entwicklungen in der Datenbankforschung werden seit Jahren in den Geo-Informationssystemen berücksichtigt; gerade modernere Produkte orientieren sich sehr stark am Stand der Forschung auf diesem Sektor. Anders sieht es auf dem Sektor der Wissensverarbeitung aus. In GIS werden bereits heute implizit - d.h. im Datenmodell und in den Programmen selbst - verschiedene Wissensformen verwendet. Sämtliche in einem GIS programmierten Abläufe müssen jedoch als Niederschreibung prozeduralen Wissens aufgefaßt werden; sie steuern das Verhalten des Systems. Wünschenswert auch im GIS-Bereich wäre allerdings die Einbeziehung anderer Wissensformen, womit sich dessen Systemarchitektur der eines Expertensystems annähert (Abbildung 12). Während die Geo-Informationssysteme bis heute noch vollständig der Klasse der konventionellen Systeme zugehören, ist durch neue Aufgabenstellungen und Integrationsbestrebungen der Wandel zu den Expertensystemen - auf jeden Fall in bestimmten Teilbereichen - unabdingbar. Die Trennung zwischen Daten und Wissen und die Abkoppelung der Verarbeitung sind wichtige Voraussetzungen für diese zukünftigen Herausforderungen.

Konventionelle Systeme



Wissensbasierte Systeme

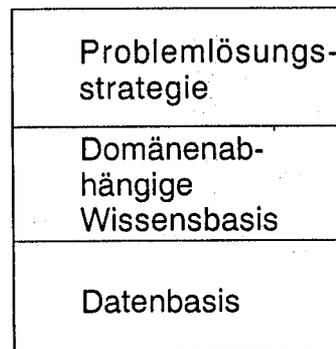


Abbildung 12 : Architektur von Programmsystemen.

5 Literaturverzeichnis

- /1/ Bill, R., Fritsch, D. (1991) : Grundlagen der Geo-Informationssysteme. Band 1 : Hardware, Software und Daten. Wichmann Verlag Karlsruhe.
- /2/ Reimer, U. (1991) : Einführung in der Wissensrepräsentation. B.G. Teubner Verlag Stuttgart.
- /3/ Bruns, W. (1990) : Künstliche Intelligenz in der Technik. Carl Hanser Verlag München Wien.
- /4/ Bill, R. (1991) : Datenbank-Technologie - Überblick und Einsatz in Geo-Informationssystemen. in : Schilcher, M. (Hrsg) : Geo-Informatik. Verlag Siemens Aktiengesellschaft Berlin und München.
- /5/ Kleine Büning, H. (1991) : Einführung in die Künstliche Intelligenz. in : Schilcher, M. (Hrsg) : Geo-Informatik. Verlag Siemens Aktiengesellschaft Berlin und München.
- /6/ Burrough, P.A. (1985) : Principles of Geographic Information Systems for Land Resources Assessment. Oxford Science Publications, Monographs on Soil and resources Survey No. 12.
- /7/ Güntsch, H. (1988) : Datenstrukturen in raumbezogenen Informationssystemen. in : X. Internationaler Kurs für Ingenieurvermessung, Dümmler Verlag Bonn.
- /8/ CODASYL (1973) : Conference on Data Systems Languages. ACM, New York.
- /9/ Codd, E.F. (1970) : A Relational Model for Large Shared Data Banks. Comm. ACM Vol. 13, No. 6, pp. 377-387.

- /10/ Schek, H.J. (1989) : Relational Database Concepts and Research Aspects to Cover Spatial Data Needs. in : Kölbl, O. (Editor, 1989) : Photogrammetry and Land Information Systems. Presses Polytechniques Romandes.
- /11/ Dittrich, K.R. (1989) : Objektorientierte Datenbanksysteme. Informatik Spektrum 12, Seite 215-218.
- /12/ Seidensticker, F.J. (1988) : HyperText : Konzepte und Beispiele. Arbeitsbericht 41, Institut für Wirtschaftsinformatik, Hochschule St. Gallen.
- /13/ Gong, J. (1990) : Object-oriented Models for Thematic Data Management in GIS. EGIS 90 Proceedings.
- /14/ Zhang, G., Tulip, J. (1990) : An Algorithm for the Avoidance of Sliver Polygons and Clusters of Points in Spatial Overlay. Proceedings of the 4th International Symposium on Spatial Data Handling. Zurich, 1990.
- /15/ Schmitgen, S. (1991) : Grundtechniken der Wissenrepräsentation und -verarbeitung. in : Schilcher, M. (Hrsg) : Geo-Informatik. Verlag Siemens Aktiengesellschaft Berlin und München.

Zusammenfassung

Der Beitrag stellte die wesentlichen Konzepte der Daten- und Wissensrepräsentation dar. Während in der Datenverwaltung in Geo-Informationssystemen die aktuelle Datenbankforschung bereits in Produkten ihren Niederschlag findet, steht die Nutzung der Konzepte der Wissensverwaltung und -verarbeitung noch aus. Mit verschiedenen Beispielen aus dem GIS-Sektor sollte die generelle Verwendbarkeit gezeigt werden.

CONCEPTS OF DATA AND KNOWLEDGE REPRESENTATION IN GEOGRAPHICAL INFORMATION SYSTEMS

Abstract

The Paper presents the concepts of data and knowledge representation. The recent database research activities are partially integrated into commercially available GIS-products. 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.

Dr.-Ing. Ralf Bill
Institut für Photogrammetrie
Universität Stuttgart
Keplerstraße 11
D 7000 Stuttgart 1