

Automatische Semantische Transformation zwischen Geo-Ontologien

BIRGIT KIELER, Hannover,

Published in: Mitteilungen des Bundesamtes für Kartographie und Geodäsie, Band 39, AGA-Tagung 2006, S. 81-90

Automatische Semantische Transformation zwischen Geo-Ontologien

(mit 8 Bildern)

Von Birgit Kieler, Hannover

ZUSAMMENFASSUNG: Damit die in großer Anzahl zur Verfügung stehenden, meist heterogenen Datensätze unterschiedlicher Herkunft zur Lösung komplexer Fragestellungen beitragen können, müssen sie mit Hilfe von Integrationstechniken kombiniert werden. Der hier vorgestellte Ansatz soll einen Beitrag zur semantischen Integration leisten, bei dem ausgehend von geometrischen Ähnlichkeiten auf semantische Entsprechungen geschlossen wird. Hierbei wird ein hoher Automatisierungsgrad angestrebt, um die derzeit hauptsächlich von Experten manuell durchgeführte Zuordnung von korrespondierenden Objektbeschreibungen in verschiedenen Datensätzen verbessern zu können. Dazu werden erste Lösungsansätze für die Ableitung semantischer Transformationsregeln vorgestellt. Dies geschieht einerseits durch die geometrische Überlagerung von Datensätzen und die daraus folgende Ableitung von Zuordnungsregeln und andererseits durch die Aufstellung von allgemeingültigen objektspezifischen Eigenschaftsfunktionen, mit deren Hilfe semantische Zuordnungen in unterschiedlichen Datensätzen möglich werden sollen.

ABSTRACT: For the purpose of ensuring that the mostly heterogeneous datasets of different origins, available in large numbers, can contribute to the solution of complex problems, they must be combined with each other by means of integration techniques. The approach presented here is intended to make a contribution to semantic integration, whereby semantic correspondences are inferred starting from geometric similarities. In this context a high degree of automation is aimed at in order to be able to improve the assignment of corresponding object descriptions in different datasets, which is presently carried out by experts mainly manually. For this purpose first solution approaches for the derivation of semantic transformation rules are presented. On the one hand this is performed by the geometric overlay of datasets and the derivation of assignment rules following therefrom, and on the other hand by the modelling of universally valid object-specific property functions, by means of which semantic assignments in different datasets shall be made possible.

RÉSUMÉ: Afin de permettre que les enregistrements de données largement hétérogènes, disponibles en grand nombre et d'origine diverse, peuvent contribuer à la solution de problèmes complexes, ils doivent être combinés à l'aide de techniques d'intégration. L'approche présentée ici doit contribuer à l'intégration sémantique, dans laquelle on conclut à des analogies sémantiques à partir de similarités géométriques. Dans ce contexte, on aspire à un haut degré d'automatisation afin de pouvoir améliorer l'allocation de descriptions d'objets correspondantes dans de différents enregistrements de données, laquelle est à l'heure actuelle réalisée par des experts surtout par voie manuelle. A cette fin, de premières approches de solution pour la dérivation de règles de transformation sémantiques sont présentées. D'un côté cela est effectué par la superposition d'enregistrements de données, et d'autre côté par l'établissement de fonctions de propriété universelles spécifiques d'objet, à l'aide desquelles des allocations sémantiques dans des enregistrements de données différents doivent être rendues possibles.

1 Einleitung und Motivation

Die Erfassung der Umwelt mit ihren Realwelt-Objekten erfolgt durch verschiedene Institutionen und ist hinsichtlich Genauigkeit und Umfang eng verknüpft mit der zu Grunde liegenden Aufgabenstellung. Die Folgen dieser anwendungsspezifischen Erfassung sind die unvermeidbare Mehrfacherfassung von Objekten und der daraus resultierende enorme Anstieg der Datenmengen, die für eine gemeinsame Nutzung eingesetzt werden sollten. Denn was nützt die große Anzahl von Daten, wenn sie bis heute aufgrund von nicht miteinander vergleichbarer Semantik nur mühsam, fehlerhaft oder gar nicht zugeordnet und sie somit auch nicht für die Lösung komplexer Fragestellungen kombiniert werden kann?

Damit wird deutlich, dass das Thema der Datenintegration noch intensiver Forschungsarbeit bedarf. Um das Ziel zu erreichen, werden grundsätzlich zwei Arten der Datenintegration unterschieden: die semantische und die geometrische Integration. Bei der semantischen Integration werden korrespondierende Objektbeschreibungen in verschiedenen Datensätzen identifiziert und für eine Zusammenführung der Daten in geeigneter Weise einander zugeordnet. Des Weiteren ist sie für den gesamten Integrationsprozess bedeutend, da sie mit der Identifizierung ähnlicher Objektarten in den verschiedenen Datensätzen erst die Voraussetzung für die geometrische Integration schafft.

Bei der geometrischen Integration wird daraufhin aus mehreren ähnlichen Datensätzen durch direkte Veränderung der Daten, beispielsweise durch die Änderung der Objektgeometrie, ein neuer, angepasster Datensatz erzeugt.

Im Rahmen dieses Beitrages werden das Forschungsvorhaben „Automatische Semantische Transformation zwischen Geo-Ontologien“ sowie die ersten Ideen zur Lösung vorgestellt. Das Projekt stellt einen Teil des deutsch-chinesischen Bündelprojekts „Interoperation of 3D Urban Geoinformation“ dar und wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert.

Das konkrete Ziel des Projektes besteht darin, automatisch semantische Übersetzungsregeln zwischen verschiedenen Geo-Ontologien – georeferenzierten Vektordatensätzen – aufzustellen und somit eine automatische Transformation zwischen ihnen herzustellen. Diese Transformation soll es ermöglichen, gleiche bzw. ähnliche Objekte in unterschiedlichen Datensätzen zu identifizieren, um Nutzern dann die daraus gewonnenen Informationen, z.B. eine aktuellere Objektgeometrie, für die Lösung von Problemstellungen zur Verfügung zu stellen. Die grundlegende Idee des Ansatzes besteht darin, ausgehend von geometrischer auf semantische Ähnlichkeit zu schließen.

Dieser Beitrag ist wie folgt gegliedert. In Kapitel 2 wird eine kurze Übersicht über auftretende Probleme und verwandte Arbeiten auf dem Gebiet der Datenintegration gegeben, bevor in Kapitel 3 auf Forschungshypothesen, den geplanten Untersuchungsablauf, Testdaten und das Testgebiet eingegangen wird. Kapitel 4 geht detaillierter auf Methoden der Trainingsphase ein und erläutert auftretende Schwierigkeiten. Eine Zusammenfassung und ein Ausblick beschließen den Beitrag.

2 Verwandte Arbeiten

In der Informatik wird nach *Gruber* (1993) eine Ontologie als eine explizite Spezifikation einer Konzeptualisierung beschrieben, wobei eine Konzeptualisierung die abstrakte und vereinfachte Sicht auf die zu repräsentierende Welt darstellt. Im Bereich der Geoinformatik kann der in Bild 1 schematisch dargestellte Objektartenkatalog des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS-OK) als ein Beispiel für eine Ontologie verstanden werden.

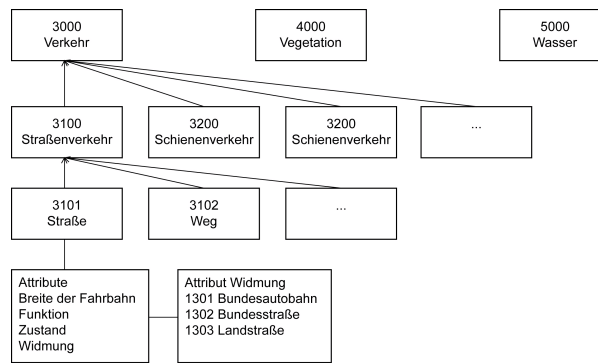


Bild 1 – Geo-Ontologie: ATKIS-Objektartenkatalog, [AdV 2006].

Bei der Datenintegration treten verschiedene Arten von Problemen auf, die es innerhalb des Integrationsprozesses zu lösen gilt (Bishr 1997). Neben technischen Problemen, die mit der Verwendung von standardisierten Datenformaten (z.B. ISO, OGC) minimiert werden können, stellt die Semantik ein sehr viel größeres Problem dar. Bei der Integration von Daten aus verschiedenen Bereichen, die durch verschiedene Institutionen bereitgestellt werden, werden Informationen zu Realwelt-Objekten sehr unterschiedlich interpretiert und dementsprechend mit unterschiedlichem, fachspezifischem Vokabular beschrieben, was die semantische Interoperabilität problematisch gestaltet. Ein weiteres Problem, das im Integrationsprozess Berücksichtigung finden muss, sind die geometrischen Differenzen, die bei der Mehrfacherfassung aufgrund unterschiedlicher Erfassungsgenauigkeiten unvermeidbar sind.

An dieser Stelle werden nur Arbeiten vorgestellt, die sich mit der Beseitigung der semantischen Heterogenität beschäftigen, da das vorgestellte Forschungsprojekt diesem Gebiet zuzuordnen ist. Kuhn (2003) prägte zur Lösung dieses Problems die Metapher von einem „Semantischem Referenzsystem“, das ähnlich dem geodätischen Referenzsystem, Transformationen zwischen verschiedenen semantischen Räumen ermöglichen soll.

Die semantische Integration, als Zuordnung semantisch ähnlicher Objektbeschreibungen, wird derzeit hauptsächlich manuell durch Experten, die über die Terminologie der erfassten Datenbestände verfügen, durchgeführt. Dadurch ist diese Arbeit sehr zeitaufwendig und kostenintensiv und kann daher nur für eine begrenzte Anzahl von Datensätzen durchgeführt werden.

Um die Probleme der semantischen Heterogenität zu lösen, werden verbreitet Ontologietechniken eingesetzt, da sie die Semantik in einer formalen Sprache repräsentieren können. Kokla (2006) stellte dazu eine Reihe von Richtlinien vor, die eine Ontologie-Integration zulassen, bei der Begriffe der individuellen Ontologie einer allgemeingültigen Ontologie zugeordnet werden.

Um den Prozess der semantischen Zuordnung zu automatisieren, werden sogenannte Instanzbasierte oder extensionale (erweiterte) Verfahren zur Bestimmung von Transformationsregeln angewendet (Kokla 2006, Duckham & Worboys 2005).

Die Motivation des Forschungsprojektes ist es, die semantischen Relationen zwischen verschiedenen Datensätzen aus rein geometrischen Eigenschaften abzuleiten, da nicht angenommen werden kann, dass z.B. die Benennung von Objekten in verschiedenen Datensätzen identisch sind. Ausgenommen davon sind eindeutig benannte Objekte, wie zum Beispiel Eigennamen von Städten oder Straßen. Diese Vorgehensweise hat Volz (2006) bereits für linienartige Datensätze durchgeführt.

3 Forschungshypothesen – Identifizierung korrespondierender Objekte

Im Rahmen des Projektes werden folgende Hypothesen aufgestellt, die das Auffinden von korrespondierenden Objektbeschreibungen ermöglichen sollen und somit etwaige Fragen, beispielsweise nach dem Vorhandensein oder der Lage von Siedlungs- oder Wasserflächen in zwei verschiedenen Datensätzen, beantworten können.

Rein geometrisch betrachtet, haben Realwelt-Objekte mit hoher Wahrscheinlichkeit etwas gemeinsam, wenn sie sich, wie in Bild 2 (links) dargestellt, in gleicher räumlicher Lage befinden und/oder ähnliche geometrische Eigenschaften aufweisen. Doch kommt es in der Praxis häufig zu geometrischen Differenzen zwischen identischen Objekten aufgrund der sehr unterschiedlichen Erfassungsgenauigkeiten, die bei der Auswertung berücksichtigt werden müssen.

Weiterhin lässt sich behaupten, dass die Identifizierung eines Objektes in zwei Datensätzen mit hoher Wahrscheinlichkeit auf eine Beziehung zwischen den beiden Beschreibungen, dargestellt in Bild 2 (rechts), schließen lässt, da es sich um das gleiche Realwelt-Objekt handelt.

Diese Hypothesen gilt es, im Rahmen der Forschungsarbeit zu beweisen und Verfahren zur Ableitung dieser Beziehungen herzuleiten.

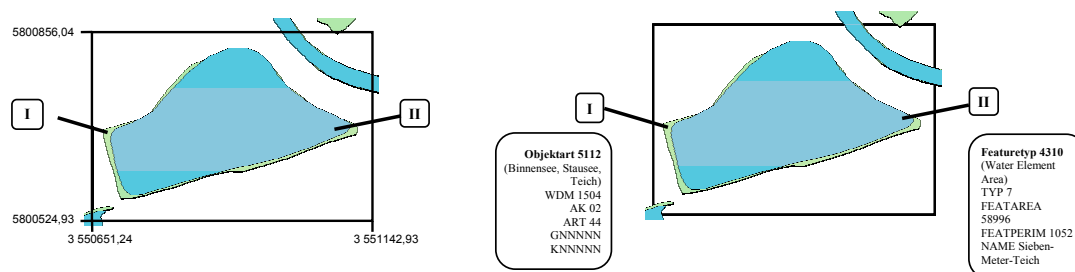


Bild 2 – Hypothesen für Gewässerobjekte: geometrisch (links), semantisch (rechts);
Datensatz I: ATKIS, Datensatz II: Teleatlas.

Für die Untersuchung der Hypothesen soll der in Bild 3 schematisch dargestellte Verfahrensablauf, der sich in zwei Phasen gliedert, umgesetzt werden. Die Trainingsphase bildet die erste Phase, die zusätzlich in eine Entwicklungs- und Validierungsphase unterteilt wird. Während in der Entwicklungsphase Datensätze des gleichen räumlichen Gebietes zum Einsatz kommen, werden in der Validierungsphase Datensätze in unterschiedlichen Gebieten verwendet. Als Ergebnis der Trainingsphase sollen für ausgewählte Objektarten Transformationsregeln TR_n abgeleitet werden, die eine Identifikation von ähnlichen Objektbeschreibungen, z.B. von Wasserflächen in verschiedenen Datensätzen, ermöglichen sollen. In der Anwendungsphase sollen dann diese aufgestellten Regeln auf einen unbekanntem Datensatz, z.B. aus China, dem Land unserer Projektpartner, angewendet werden. Der Aufbau der Trainingsphase wird in Kapitel 4 im Detail vorgestellt.

Für die in der Trainingsphase durchzuführenden Untersuchungen wird das Stadtgebiet Hannover als Testgebiet festgelegt. Die Auswahl der Testdaten wird zunächst an die Bedingung geknüpft, dass sie eine ähnliche geometrische Auflösung, d.h. den gleichen Erfassungsmaßstab besitzen. Dadurch wird sichergestellt, dass bei der Zuordnung nur Objekte mit ähnlicher räumlicher und semantischer Auflösung identifiziert werden, aus denen dann maßstabsunabhängige Regeln abgeleitet werden können. Als Geo-Ontologien wurden zu Testzwecken Daten des Amtlichen Topographisch-Kartographischen Informationssystems (ATKIS Basis DLM-Daten) und Teleatlas-Daten im GDF-Format ausgewählt.

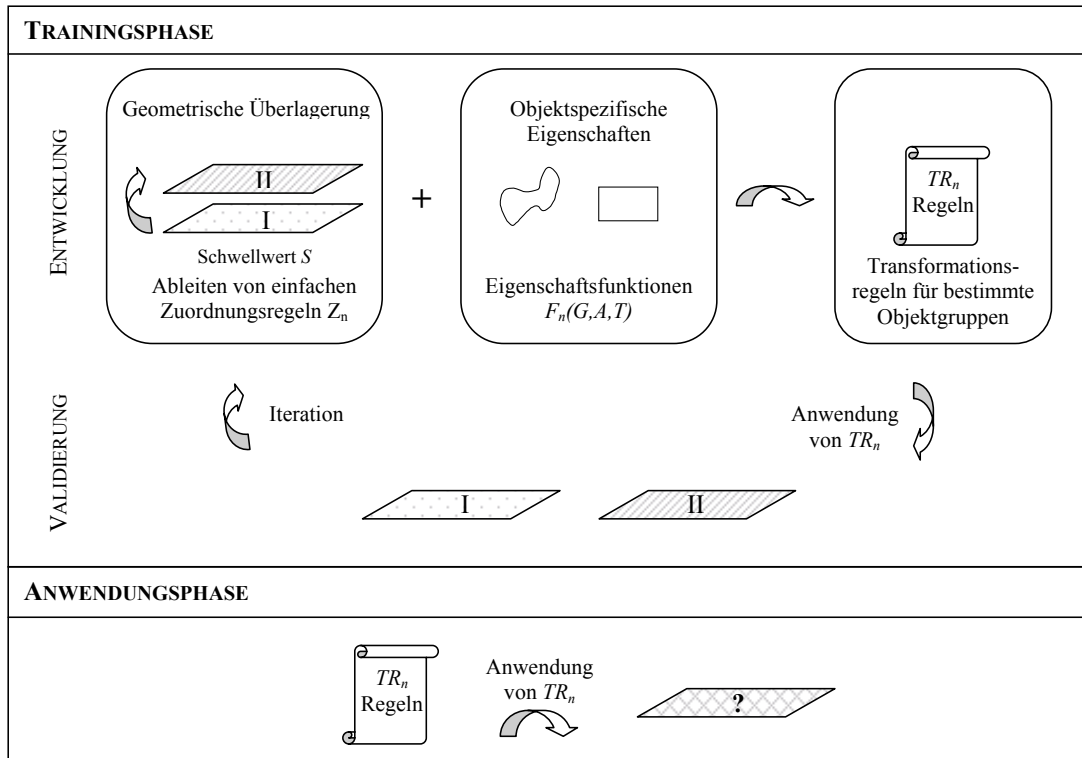


Bild 3 – Ablauf der Untersuchungsphasen: Trainingsphase (mit Entwicklung und Validierung) und Anwendungsphase.

Unabhängig von der Methode, wie ein Realwelt-Objekt in zwei verschiedenen Datensätzen identifiziert wird, ob semantisch durch die manuelle Zuordnung durch einen Experten oder geometrisch durch die automatische Überlagerung mit anschließendem Geometrievergleich, wird es im Idealfall in beiden Datensätzen gefunden. In Bild 4 ist die Einordnung eines in beiden Datensätzen gefundenen Beispielobjektes in die hier stark vereinfacht dargestellten Datenschemata gezeigt. In Datensatz I, dem das Datenmodell I zu Grunde liegt, wird das Objekt der Objektklasse A und in Datenmodell II der Objektklasse 1 zugeordnet. Wie hier zu erkennen ist, kann der Fall eintreten, dass sich die Objektklassen auf unterschiedlichen Datenebenen befinden. Die reine Zuordnung drückt aus, dass sich Objektklasse A und 1 entsprechen und sie demnach die gleiche Bedeutung haben. Doch ist das wirklich immer der Fall? Ist die Zuordnung immer eindeutig? Die Frage muss mit Nein beantwortet werden, denn es sind viele verschiedene Zuordnungen denkbar, wie z.B. dass Objektklasse A den Objektklassen 1 und 2 entspricht. Auch beschränkt sich das Problem nicht nur auf die Zuordnung der Objektklassen im Datenschema, sondern ist auch auf die Objektebene übertragbar. Die speziell auf der Objektebene auftretenden Fälle, werden in Kapitel 4 anhand von Beispielen ausführlicher betrachtet.

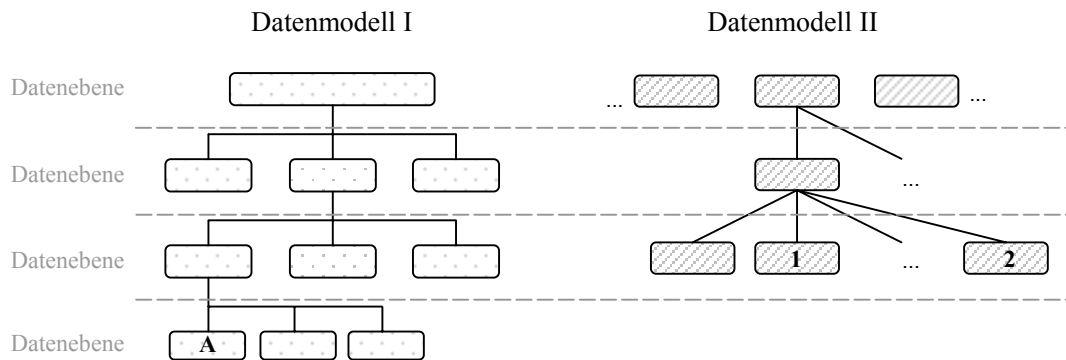


Bild 4 – Einordnung ins Datenschema; Datenmodell I: ATKIS, Datenmodell II: Teleatlas

4 Methoden der Trainingsphase und Schwierigkeiten

Für die in Kapitel 3 vorgestellten Hypothesen und die daraus abgeleiteten Untersuchungsphasen werden im Folgenden die entwickelten Lösungsstrategien und auftretenden Probleme der Trainingsphase erläutert. Ausführlich wird an dieser Stelle nur auf die Entwicklungsphase mit ihren zwei Ansätzen – der geometrischen Überlagerung und der Entwicklung von objekt-spezifischen Eigenschaften – eingegangen. Die erzielten Ergebnisse werden anschließend gemeinsam in die Ableitung von Übersetzungsregeln einfließen, die wiederum in der Validierungsphase geprüft und bewertet werden. Die Trainingsphase stellt einen iterativen Prozess dar, in dem die gesuchten Transformationsregeln solange verbessert werden können, bis die Zuordnung erfolgreich ist.

4.1 Methode der Geometrischen Überlagerung

In der Phase der Entwicklung, speziell bei der geometrischen Überlagerung, sollen identische Objekte in den ausgewählten Vektordatensätzen, die das gleiche räumliche Gebiet überdecken, durch einfache Überlagerung identifiziert werden. Aufgrund der unterschiedlichen Erfassungsgenauigkeiten muss für die eindeutige Zuordnung, bei der das Verhältnis von Schnittfläche zur Objektgröße untersucht wird, ein Schwellwert S festgelegt werden. Dieser Schwellwert soll die entstandenen, geometrischen Differenzen zwischen identischen Objekten berücksichtigen. Anhand dieser Zuordnung lässt sich wie in Bild 5 dargestellt, eine Häufigkeitsmatrix aufstellen, aus der ersichtlich wird, zu welchem Prozentsatz eine Objektklasse des einen Datensatzes einer Objektklasse des anderen zugeordnet wird.

Datensatz I	A	N_{A1}	N_{A2}	N_{A3}
	B	N_{B1}	N_{B2}	N_{B3}
	C	N_{C1}	N_{C2}	N_{C3}
		1	2	3
		Datensatz II		

Bild 5 – Häufigkeitsmatrix; N_{ij} Anzahl der eindeutig zugeordneten Objekte, $i = A, B, C$ – Objektklassen von Datensatz I und $j = 1, 2, 3$ – Objektklassen von Datensatz II.

Aufgrund dieser statistischen Analyse lassen sich so erste, einfache Zuordnungsregeln Z_n , in Form von:

Z_n (Gleichheit): Datensatz I Objektklasse A \equiv Datensatz II Objektklasse 1

Z_n (Teilmenge): Datensatz I Objektklasse A \subseteq Datensatz II Objektklasse 1

Z_n (Überlappung): Datensatz I Objektklasse A \cap Datensatz II Objektklasse 1 $\neq \emptyset$

Z_n (Disjunktheit): Datensatz I Objektklasse A \cap Datensatz II Objektklasse 1 = \emptyset

festlegen.

Bei der Aufstellung der Zuordnungsmatrix treten verschiedene Schwierigkeiten auf, die berücksichtigt und gesondert behandelt werden müssen. Allein durch die Überlagerung der beiden Datensätze wird sehr schnell die Problematik der der Realität nachempfundenen Layerstruktur offensichtlich, in der die Daten organisiert sind. Wie in Bild 6 dargestellt, kann es schon innerhalb eines Datensatzes zu folgender Situation kommen: An einer räumlichen Position befinden sich häufig mehrere Objektklassen übereinander, beispielsweise befindet sich ein See immer in einem Verwaltungsgebiet.

Infolgedessen kann das Problem auftreten, dass Objekte bereits innerhalb des eigenen Datensatzes gleichzeitig mehreren Klassen zugeordnet werden könnten. Durch das Hinzunehmen eines weiteren Datensatzes vergrößern sich die Möglichkeit und der Aufwand, die richtige Zuordnung zu finden. Dennoch kann nicht allein aus der rein geometrischen Koinzidenz auf identische Objekte geschlossen werden, die Überlagerungsflächen müssen sich auch zumindest ungefähr ähnlich sein.

Ungeachtet der Tatsache, dass innerhalb eines Datensatzes mehrere Objekte übereinander liegen, werden bereits bei der Betrachtung von jeweils nur einem Layer pro Datensatz Schwierigkeiten ersichtlich, auf die bereits in Kapitel 3 hingewiesen wurde. Dabei handelt es sich um Zuordnungsprobleme auf der Objektebene, die in Abschnitt 4.1.1 kurz an Beispielen erläutert werden.

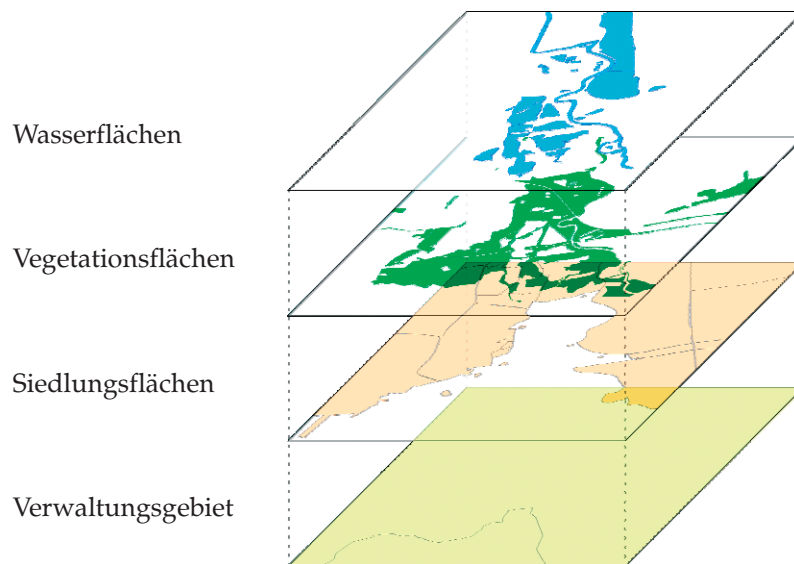


Bild 6 – Vereinfachte Darstellung der Layerstruktur am Beispiel von Datensatz I: ATKIS.

4.1.1 Zuordnungsrelationen

In Bild 7 a) bis d) sind die verschiedenen möglichen Relationen abgebildet, die im Folgenden einzeln vorgestellt werden. Es kann der Fall a) eintreten, dass zu einem Objekt kein Objekt im anderen Datensatz gefunden wird, eine sog. 1:0 Relation. Des Weiteren kann der optimale Fall b) der 1:1 Relation eintreten, so dass genau ein Objekt einem anderen zugeordnet wird. Problematisch ist dabei, dass das zugeordnete Objekt, wie in Fall c) in mehreren Objektteilen (T_1 bis T_n) modelliert wurde, die alle den gleichen Identifikator besitzen, was zu Problemen bei der Auswertung bzw. Analyse führen kann. Diesem Fall sehr ähnlich scheint Fall d) die 1:n Relation, bei dem wiederum einem Objekt mehrere eigenständige Objekte (O_1 bis O_n) zugeordnet werden müssen. Die letzte mögliche Relation ist die Zuordnung einer Gruppe von Objekten zu einer anderen Gruppe, der n:m Relation.

Um Fehlzuordnungen aufgrund der auftretenden Relationen ausschließen zu können, müssen beim Zuordnungsprozess zusätzliche geometrische Kriterien hinzugezogen werden. Welche Merkmale konkret dafür zur Verfügung stehen und wie sie zur Verbesserung der Ergebnisse beitragen können, wird im nächsten Abschnitt beschrieben.

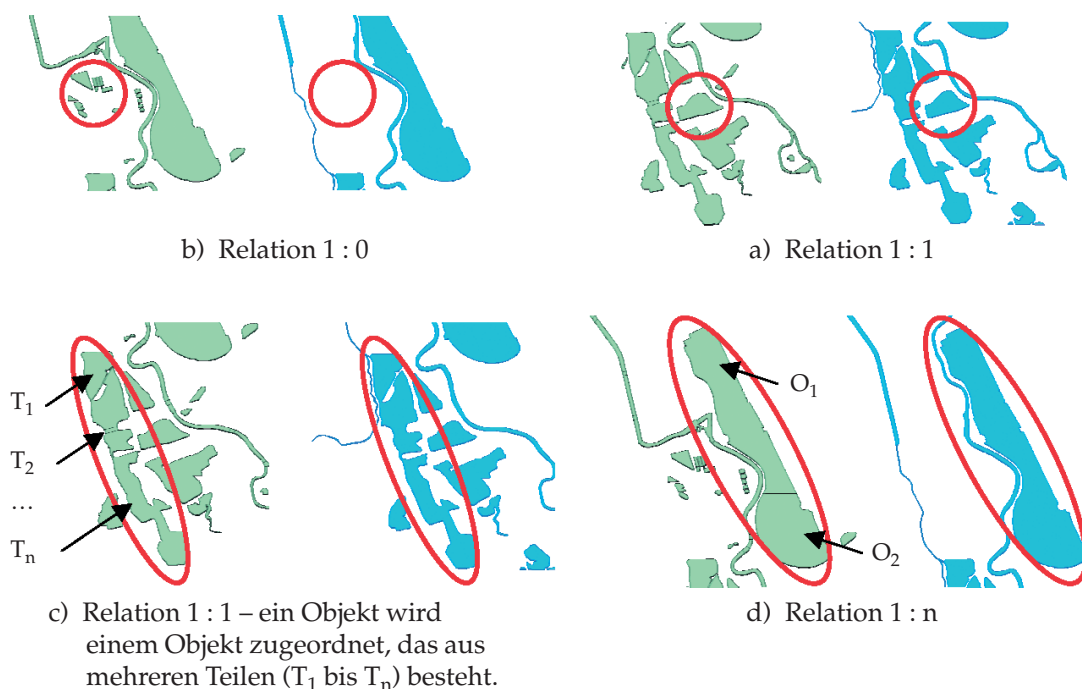


Bild 7 – Zuordnungsrelationen; Datensatz I: ATKIS (links, grün), Datensatz II: Teleatlas (rechts, blau)

4.2 Entwicklung von objektspezifischen Eigenschaften

Die einfachen Zuordnungen aus der Häufigkeitsmatrix, die als Ergebnis der geometrischen Überlagerung erzeugt werden, bieten nicht in jeder Situation zufriedenstellende allgemeingültige Transformationsregeln zwischen verschiedenen Datensätzen, da der gewählte Schwellwert S nicht optimal für alle auftretenden Situationen gewählt werden kann. Aus diesem Grund wird die Entwicklung von objektspezifischen Eigenschaftsfunktionen F_n für ausgewählte Objektarten in den Entwicklungsprozess mit aufgenommen. Neben Merkmalen

zur Objektgeometrie G sollen auch vorhandene Attributwerte A und die Topologie T zu Nachbarobjekten in den Eigenschaftsfunktionen $F_n(G, A, T)$ Berücksichtigung finden.

Eine Frage, die im Rahmen dieser Forschungsarbeit betrachtet werden muss, ist außerdem, welche formbeschreibenden Parameter zur Verfügung stehen, um die Geometrie eines Objektes präzise und objektiv beschreiben zu können, und welche außerdem das Potential besitzen, die eindeutige Identifizierung von bestimmten Objektgruppen zu ermöglichen.

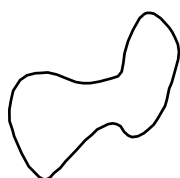
Im nächsten Abschnitt werden exemplarisch ausgewählte geometrische Merkmale vorgestellt, die diesen Anforderungen in einem bestimmten Rahmen gerecht werden können.

4.2.1 Formbeschreibende Parameter

Eines der wichtigsten geometrischen Merkmale ist der Flächeninhalt. Er bietet mit der Methode des Flächenvergleichs eine sehr einfache Möglichkeit zu prüfen, ob Objekte sich als Kandidaten eignen. Des Weiteren sind der Umfang und die Länge eines Objektes geeignet. Mit Hilfe dieser drei Grundeigenschaften lassen sich weitere Parameter berechnen, aus denen sich Rückschlüsse auf die tatsächliche Form des Objektes ziehen lassen. Dazu zählen neben der Langgestrecktheit, die das Verhältnis der beiden Hauptachsen zueinander angibt, die Kompaktheit, als Verhältnis des Umfangs und der Fläche. Außerdem geben die Rechtwinkligkeit, ein Maß der Winkeländerungen entlang der Objektkontur, sowie die Parallelität, die die Gesamtlänge der parallelen Seiten des Objektes zu dessen nichtparallelen Seiten in ein Verhältnis setzt, Aufschluss über wichtige geometrische Eigenschaften (Weindorf 2002). Auch verschiedene Arten von Hüllen, z.B. das umschließendes Rechteck minimaler Fläche oder auch die konvexe Hülle liefern eine wichtige Entscheidungshilfe für eine richtige Zuordnung.

Die Problematik besteht im nächsten Schritt darin, mit Hilfe dieser Eigenschaften eine objekt-spezifische Funktion für bestimmte Objektgruppen aufzustellen, damit eine möglichst hohe Trefferquote erfüllt wird.

Die Wahrnehmung von Objekten als Funktion dieser Eigenschaften zu modellieren, stellt eine große Herausforderung dar. In Bild 8 a) und b) soll anhand von zwei einfachen Beispielen die Schwierigkeit verdeutlicht und eine mögliche Herangehensweise zur Ableitung entsprechender Regeln vermittelt werden.



a) See, Teich



b) Gebäude

Bild 8 – Mögliche Objektarten, die es durch eine Eigenschaftsfunktion zu beschreiben gilt.

Das in Bild 8 a) dargestellte Objekt lässt sich ohne Kenntnis seiner Funktion als ein organisches Objekt ohne längere gerade Kanten beschreiben. Das Objekt in Bild 8 b) weist im Gegensatz dazu eine sehr langgestreckte, rechtwinklige Form auf. Diese Wahrnehmung der Geometrie gilt es jetzt mit Hilfe der oben vorgestellten Formparameter bzw. der Festlegung von Geltungsbereichen der Parameter zu modellieren. Für das Objekt See bzw. Teich kann die Kombination eines hohen Kompaktheitsmaßes mit einem geringen Maß für Rechtwinkligkeit eine gute Beschreibung darstellen. Beim Objekt Gebäude haben die Rechtwinkligkeit und die

Parallelität hohe Priorität. Die hier präsentierte Liste ist weder vollständig noch abgeschlossen, daher wird sie innerhalb der Entwicklungsphase gegebenenfalls ergänzt.

5 Zusammenfassung und Ausblick

In diesem Beitrag wird eine mögliche Herangehensweise zur Lösung der Problematik der semantischen Integration eingeführt. Es wird eine Methode vorgestellt, mit der aus einer Kombination von Zuordnungsrelationen und Eigenschaftsfunktionen, Transformationsregeln zwischen Datensätzen entwickelt werden können.

Die einzelnen Teilergebnisse der Entwicklungsphase sollen so miteinander kombiniert werden, dass daraus qualitativ gute Transformationsregeln ableitbar sind, die für die semantische Integration mehrerer Datensätze eingesetzt werden können. Während die aufgestellten Regeln in der Validierungsphase geprüft werden, sollen sie in der Anwendungsphase ihre Allgemeingültigkeit und Objektivität unter Beweis stellen.

Zusammenfassend lässt sich feststellen, dass für die Lösung komplexer Fragestellungen die Datenintegration zunehmend wichtiger wird. Da auf dem Gebiet der geometrischen Integration viel Forschungsarbeit geleistet worden ist und wird, muss die semantische Integration in Zukunft weiter vorangetrieben werden. Die Schwierigkeit besteht unter anderem dabei, implizites Wissen einer Nutzergemeinde, mit speziellen Kenntnissen, explizit zu machen, aber auch für die sehr subjektive Wahrnehmung von Objekten einheitliche und objektive Beschreibungsregeln zu finden.

Literatur

AdV [2006]: ATKIS-Objektartenkatalog, Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), Bonn. http://www.atkis.de/dstinfo/dstinfo2.dst_gliederung, Zugriff: 29.11.2006.

Bishr, Y., 2006: Semantic aspects of interoperable GIS. Wageningen Agricultural University and International Institute for Aerospace Survey and Earth Science (ITC), PhD thesis, Enschede.

Duckham, M.; Worboys, M. F., 2005: An algebraic approach to automated information fusion, *International Journal of Geographical Information Science*, 19 (5): 537–557.

Gruber, T. R., 1993: A translation approach to portable ontology specifications, *Knowledge Acquisition*, 5 (2): 199–220.

Kokla, M., 2006: Guidelines on geographic ontology integration, ISPRS Technical Commission II Symposium, 12–14 July, 2006, Vienna, Austria.

Kuhn, W., 2003: Semantic reference systems, in: *International Journal of Geographical Information Science*, 17 (5): 405–409.

Weindorf, M., 2002: Regelbasierte Interpretation unstrukturierter Vektorkarten, Dissertation, Deutsche Geodätische Kommission, Reihe C, Nr. 555, München.

Volz, S., 2006: Modellierung und Nutzung von Relationen zwischen Mehrfachrepräsentationen in Geo-Informationssystemen, Dissertation, Universität Stuttgart.