

Achim Lauer

Untersuchung und Bewertung zur Interoperabilität von Geodaten in räumlichen Datenbanken

Diplomarbeit

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2002 Diplomica Verlag GmbH
ISBN: 9783832464035

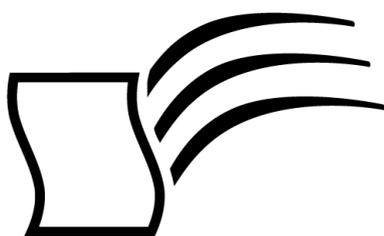
Achim Lauer

Untersuchung und Bewertung zur Interoperabilität von Geodaten in räumlichen Datenbanken

Achim Lauer

Untersuchung und Bewertung zur Interoperabilität von Geodaten in räumlichen Datenbanken

**Diplomarbeit
an der Technischen Universität Darmstadt
Fachbereich Bauingenieurwesen
September 2002 Abgabe**



Diplom.de

Diplomica GmbH _____
Hermannstal 119k _____
22119 Hamburg _____

Fon: 040 / 655 99 20 _____
Fax: 040 / 655 99 222 _____

agentur@diplom.de _____
www.diplom.de _____

ID 6403

Lauer, Achim: Untersuchung und Bewertung zur Interoperabilität von Geodaten in räumlichen Datenbanken

Hamburg: Diplomica GmbH, 2003

Zugl.: Darmstadt, Technische Universität, Diplomarbeit, 2002

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden, und die Diplomarbeiten Agentur, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

Diplomica GmbH

<http://www.diplom.de>, Hamburg 2003

Printed in Germany

Inhaltsverzeichnis

1	Einführung	4
2	Datenbanken	6
2.1	Ansätze zur Datenhaltung	6
2.1.1	Datei- oder Filesystem	6
2.1.2	Relationale Datenbank	6
2.1.3	Objektorientierte Datenbank	7
2.1.4	Objektrelationale Datenbank	8
2.2	Ansätze zur räumlichen Datenhaltung	8
2.2.1	Datei	9
2.2.2	BLOB	9
2.2.3	Relationaler Ansatz	10
2.2.4	Objektorientierter Ansatz	10
2.2.5	Objektrelationaler Ansatz	11
2.2.6	Vergleich der Ansätze	11
2.3	Indizierung	12
2.3.1	B-Baum	13
2.3.2	R-Baum	14
2.3.3	Quadbaum	17
2.3.4	Vergleich der Indizierungen	18
2.4	Informix [®] Dynamic Server 2000	20
2.4.1	Allgemeine Einführung	20
2.4.2	Installation	26
2.4.3	Administration	29
2.4.4	Schnittstellen	33
2.4.5	DataBlade Technologie	35
2.4.6	Informix Spatial DataBlade Module	36
2.4.7	Räumliche Datentypen	37
2.4.8	Räumliche Metadatentabellen	40
2.4.9	Indizierung	43

2.5	Microsoft SQL Server 2000.....	45
2.5.1	Strategie	45
2.5.2	Allgemeine Einführung.....	45
2.5.3	Editionen.....	48
2.5.4	Installation	49
2.5.5	Administration	51
2.5.6	Schnittstellen.....	53
2.5.7	Datentypen.....	54
2.5.8	Indizierung.....	55
2.6	Oracle 9i.....	56
2.6.1	Installation	56
2.6.2	Administration	56
2.6.3	Schnittstellen.....	60
2.6.4	Räumliche Datentypen.....	63
2.6.5	Räumliche Indizierung.....	64
2.7	Vergleich der Datenbanken	65
2.7.1	Vergleich zwischen Informix, SQL Server und Oracle 9i.....	65
2.7.2	Vergleich zwischen SQL Server und Access	69
3	GI-Systeme	72
3.1	MapInfo Professional 6.5.....	72
3.1.1	Anbindung an Informix	73
3.1.2	Anbindung an SQL Server.....	78
3.1.3	Anbindung an Oracle	79
3.1.4	Vergleich der Anbindungen von MapInfo.....	80
3.2	GeoMedia Professional 4.0.....	81
3.2.1	Anbindung an SQL Server.....	81
3.2.2	Anbindung an Oracle	84
3.2.3	Vergleich der Anbindungen von GeoMedia.....	85
3.3	ArcView 3.2/ArcSDE 8.1	86
3.3.1	Anbindung an Informix	92
3.3.2	Anbindung an SQL Server.....	94

3.3.3	Anbindung an Oracle	97
3.3.4	Vergleich der Anbindungen von ArcView/ArcSDE	98
4	Vergleich der Anbindung der GI-Systeme an die Datenbanken.....	100
4.1	Informix [®] Dynamic Server 2000	100
4.2	Microsoft SQL Server 2000.....	101
4.3	Oracle 9i.....	103
5	Interoperabilität.....	105
5.1	Szenarien.....	105
5.2	Interoperabilität über mehrere Datenbanksysteme	110
6	Bewertung/Fazit.....	113
7	Anhang.....	115
7.1	Fehlerbeschreibung Anbindung MapInfo an Informix	115
8	Verzeichnisse	118
8.1	Abkürzungsverzeichnis.....	118
8.2	Abbildungsverzeichnis.....	120
8.3	Tabellenverzeichnis	121
8.4	Quellenverzeichnis.....	122
8.4.1	Literatur	122
8.4.2	Internet.....	124
9	CD zur Diplomarbeit	126

1 Einführung

Geodaten, bei welchen es sich um räumliche Daten handelt, liegen heutzutage in einer Vielzahl von digitalen Vektorformaten vor. Diese können von verschiedenen GI-Systemen gelesen werden. Somit besteht die Möglichkeit, die Daten zu visualisieren, zu analysieren oder auch zu verändern. Die räumlichen Daten werden als einzelnes File auf dem Sekundärspeicher (Festplatte) gespeichert. Mittlerweile hat sich allerdings als nicht zu übersehende Alternative die Speicherung der räumlichen Daten in Datenbanken kommerzieller Anbieter entwickelt. Vorhandene GI-Systeme können nun direkt an die Datenbanken angebunden werden, um räumliche Daten zu lesen und zu schreiben. Wünschenswert wäre es nun, dass GI-Systeme von verschiedenen Anbietern auf ein und denselben räumlichen Datenbestand in der Datenbank zugreifen könnten und diesen gemeinsam visualisieren, analysieren und verändern können. Dieses gemeinsame Bearbeiten der räumlichen Daten steht in direktem Zusammenhang mit dem Begriff der Interoperabilität. Hierzu eine Definition des Begriffes:

„Interoperabilität bezeichnet die Möglichkeit, verschiedenartige Daten in einem einzelnen Arbeitsablauf zu integrieren. Dies setzt voraus, dass Syntax und Semantik der Daten dem Anwender in einheitlicher Form zur Verfügung gestellt werden. Interoperabilität erlaubt den transparenten Zugang zu mehreren raumbezogenen Daten- und Verarbeitungsressourcen innerhalb eines einzigen Arbeitsablaufes, ohne sie in einen Datenbestand zu überführen.“ ([Bil99])

Wäre eine Interoperabilität zwischen vorhandenen GI-Systemen und Datenbanksystemen möglich, so ergeben sich erhebliche Einsparungspotenziale, da räumliche Daten nicht doppelt erfasst werden müssen und auch ein wesentlich beschleunigter Workflow in vielen Bereichen des Handels, der Industrie und der öffentlichen Hand möglich wäre, da räumliche Daten zentral und ohne Informationsverluste bearbeitbar sind. Darüber hinaus kommt es zu einer erheblichen Beschleunigung des Wachstums des Geodatenmarktes. Eine Untersuchung der Ansätze und die Bewertung zur Interoperabilität von Geodaten in räumlichen Datenbanken soll in dieser Diplomarbeit erfolgen.

In Kapitel 2 werden dazu bereits bestehende Ansätze zur Speicherung räumlicher Daten in Datenbanken erläutert und verglichen. Ein Vergleich findet auch mit der bestehenden Speicherung als File auf Sekundärspeichern statt. Daran anschließend werden die durch die Speicherung in Datenbanken möglichen räumlichen Indizierungen erklärt und verglichen. Somit lassen sich Anfragen an den räumlichen Datenbestand beschleunigen. Nach diesen allgemeinen, eher theoretischen Grundlagen, werden die drei Datenbanksysteme Informix Dynamic Server, Microsoft SQL Server und Oracle 9i als Möglichkeit zur Speicherung räumlicher Daten vorgestellt. Dabei wird neben der Installation, der Administration und den Schnittstellen der Datenbank besonders auf vorhandene Ansätze zur Haltung und Indizierung von räumlichen Daten eingegangen.

Abschließend folgt ein ausführlicher Vergleich mit den Vor- und Nachteilen der drei Datenbanksysteme.

Kapitel 3 gibt eine Einführung in die drei GI-Systeme MapInfo Professional 6.5, Geomedia Professional 4.0 und ArcView3.2/ArcSDE 8.1. Es wird die Anbindungen der GI-Systeme an die Datenbanken beschrieben und analysiert. Dabei wird primär auf die Unterstützung vorhandener Ansätze zur Speicherung der räumlichen Daten in den Datenbanksystemen und deren räumlichen Indizierungen eingegangen. Bei einigen GI-Systemen erfolgen diesbezüglich auch eigene Implementierungen, die über die vorhandenen Möglichkeiten der Datenbanksysteme hinausgehen. Abschließend zu jedem GI-System erfolgt ein Vergleich zwischen den realisierten Anbindungen an die Datenbanksysteme. Kapitel 4 baut auf die im vorherigen Kapitel erläuterten Anbindungen auf. Hier werden die Anbindungen allerdings nicht von der GI-Systemseite, sondern aus der Datenbanksicht betrachtet. Zu jedem der in Kapitel 2 vorgestellten drei Datenbanksysteme erfolgt ein ausführlicher Vergleich der realisierten Anbindungen, die in Kapitel 3 vorgestellt wurden.

Kapitel 5 beschäftigt sich mit Überlegungen zur Realisierung von Interoperabilität zwischen den in Kapitel 3 beschriebenen Anbindungen. Es werden Szenarien aufgestellt und diskutiert, die es ermöglichen, räumliche Daten durch direkte Anbindung zwischen verschiedenen GI-Systemen, verschiedenen Datenbanken auszutauschen und zu bearbeiten.

In Kapitel 6 erfolgte eine abschließende Bewertung und ein Fazit zu den bestehenden Möglichkeiten zur Interoperabilität von Geodaten in räumlichen Datenbanken.

2 Datenbanken

Datenbanken haben die Aufgabe, Daten zu speichern, diese effizient zu verwalten und zu benutzen. In diesem Kapitel sollen die im Laufe der Zeit entstandenen verschiedenen Ansätze zur Realisierung einer Datenbank erläutert werden. Darauf aufbauend erfolgte Umsetzungen zur Speicherung von räumlichen Daten. Anschließend folgen die auch für räumliche Daten wichtigen Datenbankobjekte Indizes. Des Weiteren werden die drei Datenbankprodukte Informix, SQL Server und Oracle vorgestellt und abschließend erfolgt ein Vergleich derselben.

2.1 Ansätze zur Datenhaltung

Allgemein wird zwischen sechs fundamentalen Ansätzen unterschieden, um Daten zu halten bzw. zu verwalten:

- Datei- oder Filesystem
- Hierarchische Datenbank
- Netzwerkdatenbank
- Relationale Datenbank
- Objektorientierte Datenbank
- Objektrelationale Datenbank

Jeder Ansatz hat seine eigenen Stärken, die ihn somit für bestimmte Datenbankaufgaben zweckmäßig erscheinen lässt. Die aufgeführten Ansätze werden, mit Ausnahme der hierarchischen Datenbank und Netzwerkdatenbank, da diese zur Speicherung von räumlichen Daten nicht von Interesse sind, im folgenden beschrieben.

2.1.1 Datei- oder Filesystem

Datei- oder Filesysteme kamen Anfang der 60er Jahre auf. Die Daten wurden in elementaren Dateien abgelegt und die Verwaltung der Daten übernahm direkt die Anwendung. Somit ergaben sich verschiedene Nachteile. Die Datenorganisation war geräteabhängig und zwangsweise redundant, da nicht alle Daten zentral abgespeichert wurden und somit entstanden leicht inkonsistente Datenbestände. Siehe hierzu auch die Abb. 2-1 auf Seite 7.

2.1.2 Relationale Datenbank

Nach den Dateisystemen kamen die relationalen Datenbanksysteme auf, welche derzeit auf dem Markt der Datenbanken den größten Marktanteil besitzen. Als Beispiele seien Oracle, DB2, Informix, Microsoft SQL Server und Sybase genannt.

Die Daten werden bei relationalen Datenbanken mengenorientiert in Form einer Tabellenstruktur, in der alle Datentypen und Beziehungen definiert werden, in Tabellen (Relationen) gespeichert. Für eine ausführliche Einleitung in Relationen siehe [Hei01]. Kennzeichen dieser Datenbanken ist, dass sie eine Drei-Ebenen-Architektur nach dem ANSI-SPARC Modell besitzen. Dadurch ist eine strikte Trennung von Daten (auf der Festplatte), Datenlogik (Verwaltung übernimmt das Datenbankmanagementsystem (DBMS)) und Anwendungslogik (z. B. GIS-Client) gewährt. Außerdem existiert eine einheitliche Datenbanksprache zur Definition, Abfrage und Manipulation der Daten (z. B. SQL), eine Einbettung dieser Sprache in kommerzielle Programmiersprachen (SQLJ), kontrollierter Mehrbenutzerbetrieb sowie Datensicherheitsmechanismen.

Relationale Datenbanken sind optimiert auf schnelle, kurzlaufende Abfragen und Transaktionen für folgende einfache Datentypen:

- Integer
- Float
- Character, String
- Datum und Uhrzeit, Zeitintervall

Einige relationale Datenbanken besitzen darüber hinaus noch eine beschränkte Unterstützung für komplexe Datentypen. Diese werden in sogenannten „simple large objects“ gespeichert. Dies sind reine Text- bzw. Byte-Typen, bei denen weder eine Indizierung, Suchalgorithmen noch Manipulationen durch das Datenbankmanagementsystem (DBMS) möglich sind. Sie sind ähnlich denen in 2.2.2 aufgeführten BLOB und CLOB.

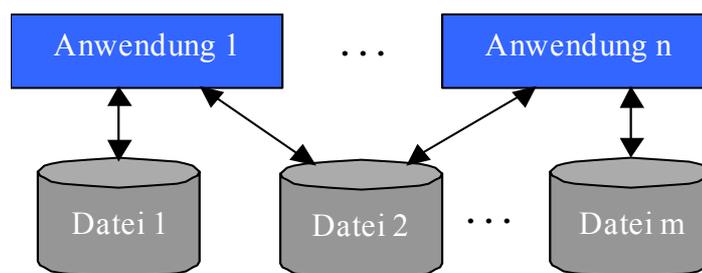


Abb. 2-1: Datei- oder Filesystem

2.1.3 Objektorientierte Datenbank

Objektorientierte Datenbankmanagementsysteme sind kommerziell seit 1987 verfügbar. Ihr Marktanteil ist im Vergleich zu relationalen Datenbanken verschwindend gering. Entstanden sind diese aus einer Zusammenführung der Konzepte von Datenbanken und objektorientierten Programmiersprachen (z. B. Java, C#, C++).

Eine objektorientierte Datenbank erweitert die klassische relationale Datenbank um komplexe Datentypen, die mit Typkonstruktoren wie *set of*, *tuple of* und *list of*

beschrieben werden können. Außerdem bekommt jedes Objekt eine Objektidentität (object identifier). Somit ist es möglich vorhandene Objekte von ihren Werten, die sie besitzen, zu unterscheiden. Letztendlich ist auch eine Vererbung der Attribute der existierenden Objekte möglich. Manche objektorientierte Datenbanken beherrschen auch höhere objektorientierte Konzepte wie z. B. Metaklassen (in Java mit abstrakten Klassen bezeichnet), Methoden, die auf Objekten und ihren Attributen operieren, Vererbung und Überschreiben von Methoden sowie Kapselung.

2.1.4 Objektrelationale Datenbank

Objektrelationale Datenbanken sind der neueste Ansatz zur Haltung von großen Datenbeständen. Sie versuchen die Vorteile von relationalen und objektorientierten Datenbanken zu vereinigen sowie die Nachteile weitestgehend auszuschalten. Daher bieten sie fortgeschrittene Unterstützung für komplexe Datentypen, die in sogenannten „smart large objects“ gespeichert werden. Diese werden auch BLOB (Binary Large Object) bzw. CLOB (Character Large Object) genannt.

Darüber hinaus erlauben sie die Definition von eigenen Datentypen als Kombination aus einfachen Datentypen, bereits definierten Datentypen. Diese eigenen Datentypen werden „User Defined Type“ (UDT) genannt. UDT's können auch durch Vererbung und Ergänzung bereits bestehender UDT's erzeugt werden. Außerdem ist es möglich, eigene Funktionen zu entwickeln, die auf diesen UDT's arbeiten. Diese Funktionen sind mit Methoden aus objektorientierten Programmiersprachen vergleichbar. Es gibt z. B. spezielle UDT's für die Arbeit mit 2D und 3D-Bildern, Musik, Video, elektronische Dokumente, HTML-Seiten und räumlichen Daten. Auf UDT's für räumliche Daten wird sowohl in 2.4.7 als auch in 2.6.4 noch genauer eingegangen. Ebenfalls bekommen instanzierte Objekttypen vom DBMS eine systemweite eindeutige Objektidentität zugewiesen. Insgesamt gesehen ermöglichen objektrelationale Datenbanksysteme bessere Speicher- und Manipulationsmöglichkeiten für komplexe Datenstrukturen als relationale Datenbanksysteme.

([Heu00],[Inf99a],[Orf99])

2.2 Ansätze zur räumlichen Datenhaltung

Nachdem die einzelnen Ansätze zur Haltung von Daten erläutert wurden, werden nun darauf aufbauend verschiedene Umsetzungen zur Speicherung von räumlichen Daten beschrieben. Diese Datenbanken werden dann räumliche Datenbanken genannt. Unter räumlichen Daten werden Geometrien verstanden, welche in der Regel in einer Vektorform vorliegen. Oftmals sind mit diesen noch Sachattribute verknüpft, welche gemeinsam gespeichert werden können. Bei sämtlichen nun folgenden Umsetzungen wird davon ausgegangen, dass die Geometrien in Vektorform vorliegen. Dies hat einige Vorteile im Vergleich zu einem Rasterformat, u. a. erweiterte Abfragemöglichkeiten, höhere Datenkompressionen, leichte Überführung in andere Koordinatensysteme.