

## **Diplomarbeit**

im Fachgebiet Wirtschaftsinformatik  
am Lehrstuhl für Wirtschaftsinformatik und Interorganisationssysteme

### **Repräsentations- und Abfragemechanismen für geographische Informationen im Web**

Themensteller: Prof. Dr. Stefan Klein

Betreuer: Dr. Bernd Schneider

vorgelegt von: Angela Schwering  
Steinfurter Str. 77  
48149 Münster  
0251-83875060  
wiansc@wi.uni-muenster.de

Abgabetermin: 2002-12-20

## Inhaltsverzeichnis

Inhaltsverzeichnis .....	II
Abbildungsverzeichnis .....	III
Abkürzungsverzeichnis.....	V
1 Einleitung.....	1
2 Zentrale Begriffe .....	4
2.1 Geographische Information.....	4
2.2 Suchmaschinen .....	6
3 Anwendungsszenarios.....	7
3.1 Endkunden-orientierte Anwendung.....	9
3.2 Business-orientierte Anwendung.....	13
3.3 Administrations-orientierte Anwendung .....	15
4 Zusammenstellung und Diskussion konzeptioneller Ansätze.....	20
4.1 Informationserhebung im Web.....	21
4.2 Repräsentation im Web .....	24
4.3 Informationsakquisitionskomponente der Suchmaschine.....	31
4.4 Informationsbasis der Suchmaschine .....	37
4.5 Abfragekomponente der Suchmaschine .....	38
4.6 Präsentationskomponente der Suchmaschine .....	44
4.7 Offene Fragen .....	47
5 Technischer Ansatz.....	49
5.1 META-Tag in HTML .....	49
5.2 Resource Description Framework (RDF).....	51
5.2.1 RDF Data Model.....	52
5.2.2 Realisierung.....	52
5.2.3 Namespace für geographische Elemente.....	53
5.2.4 Namespace für textuelle Elemente.....	59
5.2.5 Lösungsvorschlag in XML-Syntax .....	61
5.3 Resource Description Framework in HTML.....	63
5.4 Diskussion.....	65
6 Zusammenfassung und Ausblick.....	67
Literaturverzeichnis .....	69
Anhang .....	72

## Abbildungsverzeichnis

Abb. 3.1:	Erster Teil des Morphologischen Kastens .....	7
Abb. 3.2:	Zweiter Teil des Morphologischen Kastens .....	8
Abb. 3.3:	Suchablauf einer Endkunden-orientierten Anwendung [1].....	10
Abb. 3.4:	Suchablauf einer Endkunden-orientierten Anwendung [2].....	11
Abb. 3.5:	Analyse der Endkunden-orientierten Anwendung.....	12
Abb. 3.6:	Ergebnisdarstellung einer Business-orientierten Anwendung.....	13
Abb. 3.7:	Analyse der Business-orientierten Anwendung.....	15
Abb. 3.8:	Suchanfrage einer Aministrations-orientierten Anwendung.....	16
Abb. 3.9:	Ergebnisdarstellung einer Aministrations-orientierten Anwendung .....	17
Abb. 3.10:	Analyse der Aministrations-orientierten Anwendung .....	18
Abb. 4.1:	Struktur einer geographischen Suche .....	20
Abb. 4.2:	Informationserhebung und Repräsentation im Internet.....	25
Abb. 4.3:	Kontaktadresse in HTML.....	27
Abb. 4.4:	Verschiedene Kontaktadressen.....	28
Abb. 4.5:	Kontaktadresse innerhalb eines <address>-Tags .....	29
Abb. 4.6:	Informationserhebung im Internet und -akquisition der Suchmaschine.....	32
Abb. 4.7:	Informationen eines WWW-Whois-Dienstes .....	34
Abb. 4.8:	Informationen eines IP-Whois-Dienstes.....	34
Abb. 4.9:	Informationen der Localizer-Datenbank.....	35
Abb. 4.10:	Raumbezogene Suchanfragen .....	39
Abb. 4.11:	Räumliche Datentypen .....	40
Abb. 4.12:	Räumliche Operatoren I .....	41
Abb. 4.13:	Räumliche Operatoren II.....	41
Abb. 4.14:	Räumliche Operatoren III.....	41
Abb. 4.15:	Vergleich und Bewertung der Konzepte zur Informationserhebung.....	48
Abb. 5.1:	Das META-Element.....	49
Abb. 5.2:	META-Tag "geo.position" .....	50
Abb. 5.3:	META-Tag-Codierung durch Koordinaten .....	50
Abb. 5.4:	META-Tag-Codierung durch eine Adresse.....	51
Abb. 5.5:	Beispiel für ein Statement .....	52
Abb. 5.6:	Graphische Darstellung eines Statements.....	53
Abb. 5.7:	Graphische Darstellung eines geographischen Bezugs in RDF .....	53
Abb. 5.8:	Die drei Schemata von GML.....	54
Abb. 5.9:	Geographische Basistypen .....	54
Abb. 5.10:	Schema für einen Point in GML.....	55
Abb. 5.11:	Beispielcode für einen Point in GML.....	55
Abb. 5.12:	Schema für einen LineString in GML .....	56
Abb. 5.13:	Beispielcode für einen LineString in GML .....	56
Abb. 5.14:	Schema für ein Polygon in GML [1] .....	57
Abb. 5.15:	Beispiel für ein Polygon.....	57
Abb. 5.16:	Schema für ein Polygon in GML [2].....	57
Abb. 5.17:	Beispielcode für ein Polygon in GML.....	58
Abb. 5.18:	Schema für einen MultiPoint in GML.....	58
Abb. 5.19:	Beispielcode für einen MultiPoint in GML .....	59
Abb. 5.20:	Schema für eine Adresse [1] .....	59
Abb. 5.21:	Beispielcode für eine Adresse .....	60
Abb. 5.22:	Schema für eine Adresse [2] .....	61
Abb. 5.23:	Graphische Darstellung einer RDF-Codierung mit GML Point.....	61

Abb. 5.24: RDF-Codierung mit GML Point in XML.....	62
Abb. 5.25: Graphische Darstellung einer RDF-Codierung mit Adresse.....	62
Abb. 5.26: RDF-Codierung durch eine Adresse in XML.....	63
Abb. 5.27: RDF-Codierung mit Koordinaten in HTML [1] .....	64
Abb. 5.28: Interpretation des RDF-Codes in einem Browser .....	64
Abb. 5.29: RDF-Codierung mit Koordinaten in HTML [2] .....	65
Abb. 5.30: Aufwand-Nutzen-Relation .....	65

## Abkürzungsverzeichnis

BLOB	Binary Large Object
CAIDA	Cooperative Association for Internet Data Analysis
DBMS	Database Management System
DNS	Domain Name System
DNS LOC	Domain Name System Locator
Geodaten	geographische Daten
Geoinformation	geographische Information
GIF	Graphics Interchange Format
GIS	Geodateninformationssystem
GML	Geography Markup Language
GPS	Global Positioning System
GPRS	General Packet Radio Service
HTML	Hypertext Markup Language
IP	Internet Protocol
ISO	International Organization for Standardization
JPEG, JPG	Joint Photographic Expert Group
MP3	MPEG Audio Layer III
MPEG	Moving Picture Expert Group
OGC	Open GIS Consortium
PDF	Portable Document Format
RDF	Resource Description Framework
SQL	Structured Query Language
UMTS	Universal Mobile Telecommunications System
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	Extended Markup Language

## 1 Einleitung

Die Menge an Daten und Informationen im Internet ist so groß, dass es häufig unmöglich ist, durch gezieltes Ausprobieren von Internet-Adressen die gesuchten Informationen zu erhalten. Um die Masse an veröffentlichten Informationen nutzbar zu machen, ist ein Werkzeug zur Unterstützung einer umfassenden und systematischen Informationssuche unabdingbar. Dies führte schon früh zur Entwicklung sogenannter Suchmaschinen. Diesen Suchmaschinen kommt – insbesondere vor dem Hintergrund der immer weiter steigenden Informationsmenge – eine herausragende Bedeutung zu, die sich beispielsweise in neuen und verbesserten Funktionen der Suchmaschinen niederschlägt.

Derzeit beschränken sich die in Suchmaschinen eingesetzten Strategien auf das Konzept der textuellen oder kategoriellen Suche. Eine *textuelle Suche* gibt zu einem Suchbegriff oder einer komplexen, zusammengesetzten Suchanfrage alle Ressourcen zurück, die in irgendeiner Form die Suchanfrage im Text enthalten. Diese Suchergebnisse werden mittels eines suchmaschinenspezifischen Algorithmus sortiert und dem Benutzer als Liste anklickbarer Links präsentiert. Diese Art der Suche birgt einige Probleme auf syntaktischer und semantischer Ebene. Aus falscher Schreibweise oder bei Veränderungen im Wortstamm kann eine schlechte Trefferquote resultieren. Homonyme und Synonyme werden durch Suchmaschinen oft nur unzureichend erkannt, genauso wie Verneinungen.

Bei einer *kategoriellen Suche* wird versucht, das gesamte Wissen im Web durch Ober- und Unterkategorien bzw. durch ein Netzwerk von Kategorien zu organisieren und zu verwalten. Die Kategorien und die Beziehungen dazwischen sind das eigentliche Know-how der Suchmaschinen. Der Suchweg ist jedoch auf die bestehenden und von der Suchmaschine angebotenen Kategorien beschränkt. Anfragen, die sich schlecht einer bestimmten Kategorie zuordnen lassen, liefern meist unzureichende Ergebnisse.

Um die vorstehend angesprochenen Schwachstellen existierender Suchmaschinen abzuschwächen oder gar zu beseitigen, müssen andersartige Kriterien oder Strukturierungen für eine neue Art der Suche herangezogen werden.

Ca. 80% aller Daten und Informationen, mit denen der Mensch tagtäglich konfrontiert wird, haben einen Raumbezug. Sich in einer dreidimensionalen Welt zurechtzufinden gehört zu den gewohnten Aufgaben eines Menschen. Daher nutzt der Mensch häufig diesen räumlichen Bezug, um sich Daten und Informationen zu merken, miteinander in Beziehung zu setzen und zu verarbeiten. Die oben beschriebenen Suchstrategien kommen den natürlichen Fähigkeiten des Menschen, Informationen über seine raumbezogene Denkweise zu speichern und zu finden, nicht entgegen.

Es ist daher sinnvoll, die Funktionalität von Suchmaschinen neben der derzeit realisierten textuellen und kategoriellen Suche auf die Ebene der räumlichen oder *geographischen Suche* zu erweitern. Statt auf Syntax und Kategorien basiert die geographische Suche auf räumlichen Bezügen wie Koordinaten, aber auch auf textuellen Angaben wie Adressen oder Ortsnamen. Diese Informationen müssen auf Internet-Sites mit geographischem Bezug enthalten bzw. aus diesen extrahierbar sein. Zusätzlich kann von Dritten, z.B. Dienstleistern oder Intermediären, ein "Geo-Coder" angeboten werden, der eine Übersetzung von Sachnamen oder textuellen Ortsangaben in geographische Referenzen leistet. So kann eine geographische Suche als alternative oder ergänzende Such- und Navigationsform angewendet werden.

Im Gegensatz zur textbasierten Suche ist die geographische Suche eine White Box-Suche. Während bei der textuellen Suche ein suchmaschinenspezifischer, dem Nutzer meist unbekannter Suchalgorithmus ausgeführt wird, stellt die Karte bei der geographischen Suche für die Visualisierung der Ergebnisse einen absoluten Maßstab dar. So kann der Nutzer den Zusammenhang zwischen seinem geographischen Suchbegriff und dem Ergebnis unmittelbar nachvollziehen und das Suchgebiet bewusst beeinflussen.

GPRS und in Zukunft auch UMTS können die Bandbreite zur Verfügung stellen, solche Anwendungen auch mobil zu nutzen. Gerade hier wird ein großes Anwendungspotenzial vermutet, da die eigene geographische Position häufig Ausgangspunkt einer Suche sein dürfte. Der Einsatz geographischer Daten im Web ermöglicht die Entwicklung neuartiger Suchstrategien und Navigationsformen.

In der vorliegenden Arbeit wird untersucht, wie sich eine raumbezogene Strukturierung auf die Informationssuche im Web anwenden lässt. Zunächst werden in Kapitel zwei zentrale Begriffe, die für die Suche nach geographischen Daten von Bedeutung sind, erläutert. Dazu werden Geoinformationen eingeführt und ein Überblick über die verschiedenen Arten von Suchmaschinen gegeben. Durch Anwendungsszenarios wird im dritten Kapitel die praktische Relevanz und das Anwendungspotenzial für die Praxis dargestellt.

In Kapitel vier werden die Schritte zur Realisierung einer geographischen Suchmaschine beschrieben. Diese Struktur erfordert es vorab, die Voraussetzung für eine geographische Suche im Web erst zu schaffen. Zunächst muss der räumliche Bezug einer Web-Site identifiziert und maschinenlesbar dargestellt werden. Dazu wird eine Repräsentationsform erarbeitet, die geeignet ist, einen hohen Verbreitungsgrad zu erreichen. Danach wird gezeigt, wie durch Komponenten der Suchmaschine eine solche Suchanfrage bearbeitet werden kann. Nach einer abschließenden Bewertung wird festgestellt,

dass in der Informationserhebung noch Defizite bestehen, speziell in der Modellierung bzw. Repräsentation des geographischen Bezugs.

In Kapitel fünf werden zur Überwindung der Defizite drei alternative Konzepte zur Informationserhebung entwickelt. Sie basieren alle auf der Beschreibung des geographischen Bezugs einer Web-Site durch Metadaten, unterscheiden sich jedoch durch die Form der Codierung. Für diese Geodaten werden Referenzschemata entwickelt, in denen bestehende, weit verbreitete Sprachen und Konzepte aus verschiedenen Gebieten miteinander kombiniert und durch eigene Erweiterungen ergänzt werden. Die resultierenden Repräsentationen eignen sich für verschiedenste Einsatzzwecke.

## 2 Zentrale Begriffe

Im Folgenden werden die für die vorliegende Arbeit relevanten Begriffe eingeführt. Es wird erklärt, was unter geographischen Informationen zu verstehen ist, und ein Überblick über Suchmaschinen gegeben.

### 2.1 Geographische Information

"Information bezeichnet ein explizites Wissen, das Menschen nutzen oder bereitstellen"<sup>1</sup>. Durch die "Explizitheit [...] wird gefordert, dass Informationen unabhängig vom Menschen verfügbar sind oder verfügbar gemacht werden können"<sup>2</sup>. Unter Information wird auch Wissen, „das jemand beim Handeln zur Erreichung eines angestrebten Ziels benötigt und das durch die Anwendung von Regeln und Anweisungen auf Daten entsteht“<sup>3</sup>, verstanden.

Unter Geoinformation wird Information mit einem Raumbezug verstanden. Die verstärkten Forderungen, Information als separaten Produktionsfaktor zu behandeln, weisen auf ihre zunehmende Bedeutung hin.<sup>4</sup> Diese stellt neben den drei Produktionsfaktoren Arbeit, Boden, Kapital eine zentrale Ressource in der heutigen Welt dar. Um die speziellen Eigenschaften von Geoinformationen zu verdeutlichen, werden im Folgenden markante Charakteristika aufgeführt:

- Das wohl wichtigste Merkmal der Geoinformation ist ihr Raumbezug. Dieser wird beispielsweise durch Koordinaten, Adressen oder andere Raumbezugsformen dargestellt.
- Geoinformationen lassen sich teilen und zusammenfassen und können so auf unterschiedlichen Aggregationsstufen genutzt werden. Durch Kombinieren und Erweitern der Geoinformationen, z.B. durch Zusammenführen von Straßen- und Gewässerkarten, werden neue Informationen generiert.
- Geoinformation als digitales Produkt ist quasi unbegrenzt transportfähig. Das Internet als globales Netzwerk wird immer häufiger als Transportmedium genutzt.

---

<sup>1</sup> Teubner, R. (1997), S. 17.

<sup>2</sup> Teubner, R. (1997), S. 17.

<sup>3</sup> Bill, R.; Zehner, M. (2002), S. 137.

<sup>4</sup> Vgl. Seidenberg, U. (1998), S. 3-5.

- Durch Geoinformation wandeln sich Arbeitsweise und Anforderungen in vielen Wirtschaftszweigen. Durch Information können häufig Arbeitsabläufe optimiert und Arbeitskraft ersetzt werden.<sup>5</sup>

Informationen werden durch Daten repräsentiert, deren spezifische Form sich für die technische Be-/Verarbeitung eignet. "Unter Daten werden Zeichen (Symbole) verstanden, die aufgrund bekannter oder unterstellter Abmachungen Informationen zum Zwecke der Verarbeitung darstellen"<sup>6</sup>. Um Daten (voll-)automatisch durch Computer verarbeiten zu können, müssen sie in formatierter Form vorliegen.

Geoinformationen werden durch Geodaten - „Daten über Gegenstände, Geländeformen und Infrastrukturen an der Erdoberfläche“<sup>7</sup> - dargestellt. Als wesentliches Merkmal muss auch hier der Raumbezug vorliegen.<sup>8</sup> Geodaten setzen sich aus verschiedenen Daten zusammen:

- Geometriedaten wie beispielsweise die Lage und die Form der Objekte, die beschrieben werden,
- Topologiedaten, d.h. explizit gespeicherte Daten über die Lage und Anordnung geometrischer Gebilde im Raum und deren Beziehungen,
- graphische Ausprägungen der beschriebenen Objekte, z.B. die Signaturen, deren Farbe und Typographie,
- Sachdaten wie beispielsweise alphanumerische Daten zur Beschreibung der Semantik.<sup>9</sup>

Wegen des hohen Erfassungsaufwandes für Geodaten stellt Geoinformation ein besonders wertvolles Wirtschaftsgut dar. Durch die großen Datenmengen entstehen für die Verarbeitung von Geoinformationen typische Probleme bei der Realisierung akzeptabler Antwortzeiten beim Zugriff und bei der Verwaltung der außerordentlich komplexen Beziehungen der Objekte untereinander.

---

<sup>5</sup> Vgl. Bill, R.; Zehner, M. (2001), S. 110.

<sup>6</sup> Teubner, R. (1997), S. 18.

<sup>7</sup> Bill, R.; Zehner, M. (2001), S. 106.

<sup>8</sup> Vgl. Bill, R.; Zehner, M. (2001), S. 106-107.

<sup>9</sup> Vgl. Bill, R.; Zehner, M. (2001), S. 106-107.

## 2.2 Suchmaschinen

Eine Suchmaschine muss am Front-End die Eingabe der Anfrage und die Präsentation der Suchergebnisse realisieren. Im Back-End findet die Indizierung der im Web verfügbaren Ressourcen, das Matching von Anfrage und Ressourcen und die Sortierung der Ergebnisse statt.

Bei einer Benutzeranfrage werden anhand einer speziellen Abfragesprache Informationen aus der Datenbasis der Suchmaschine abgefragt.<sup>10</sup> Variationsmöglichkeiten der Suchanfrage wie Anzahl und Kombinationsmöglichkeiten von Suchbegriffen, Sprach- und Zeitvorgaben sind ein weiteres Unterscheidungskriterium von Suchmaschinen.

Nachdem die der Anfrage entsprechenden Suchergebnisse gefunden sind, muss die Trefferliste durch ein Relevanz-Ranking sortiert werden. Hier werden von Suchmaschinen individuelle, komplizierte Algorithmen verwendet, die die Worthäufigkeit, die Wortabstände und -dichte und die Spezifität der Begriffe unterschiedlich bewerten. Die Ergebnisse werden dann mit URL-Adresse und Kurzbeschreibung aufgelistet und dem Anfragenden präsentiert.<sup>11</sup>

Hinsichtlich der Indizierung der Ressourcen können zwei Typologien von Suchmaschinen unterschieden werden:

Verzeichnisbasierte Suchdienste wie Kataloge bzw. Directories basieren auf einer manuellen Datenerhebung. Web-Site-Autoren oder Lektoren tragen interessante Webseiten in die Datenbasis der Suchdienste ein. Kataloge können anhand von verschiedenen Kriterien, z.B. Sachgebieten, durchsucht werden.<sup>12</sup>

Automatische Datenakquisition kann durch roboterbasierte Suchdienste realisiert werden. Diese arbeiten sich anhand der Hypertextverweise durch das Netz, führen eine lexikalische Analyse durch, extrahieren inhaltsrelevante Terme aus dem Dokument und speichern diese in einer Datenbasis. Verschiedene lexikographische Untersuchungen zur Wortstamm-Isolierung werden auf dem Datenbestand durchgeführt und Beziehungen zwischen Begriffen mit einem Thesaurus systematisch erkannt. So wird ein Index zu den WWW-Dokumenten erstellt.<sup>13</sup>

---

<sup>10</sup> Vgl. Bill, R. (1996), S. 346.

<sup>11</sup> Vgl. Teuteberg, F. (1997), S. 377-380.

<sup>12</sup> Vgl. Teuteberg, F. (1997), S. 373.

Vgl. Steinberg, S. (1996).

<sup>13</sup> Vgl. Mertens, P. (2001), S. 446.

Vgl. Teuteberg, F. (1997), S. 374-375.

Vgl. Chang, E. (1996).

### 3 Anwendungsszenarios

Im Folgenden werden Anwendungsszenarios aus dem Endkunden-, dem Business- und dem administrativen Bereich vorgestellt, um verschiedene Einsatzmöglichkeiten einer geographischen Suche aufzuzeigen und die Relevanz für die Praxis zu belegen. Analyse und Bewertung werden anschließend in einem Morphologischen Kasten zusammengefasst.

#### Morphologischer Kasten

Zunächst wird untersucht, wie geographische Informationen Dokumenten zuzuordnen sind und wie die Zuordnung in der Suchmaschine gespeichert wird. Der Datenbestand muss effizient durchsucht und nach vorgegebenen Kriterien abgefragt werden können (s. Abb. 3.1).

Informationserhebung	direkt	indirekt	automatisch
Datenpflege	zentral		dezentral
Aktualisierung des Geo-Bezugs	manuell	Aktivierung manuell, Durchführung automatisch	automatisch
Fehlerquote der Suchmaschine	gering		hoch
Anwendungsbereich	Internet		Intranet
Skalierbarkeit	hoch		beschränkt

Abb. 3.1: Erster Teil des Morphologischen Kastens

Unter Informationserhebung wird die Feststellung des geographischen Bezugs und seine Veröffentlichung verstanden. Während die direkte Informationserhebung durch den Web-Site-Autor geschieht, sind an dem Prozess der indirekten Erhebung Dritte beteiligt. Die automatische Informationserhebung wird von Maschinen durchgeführt.

Die Pflege der veröffentlichten Geoinformationen kann entweder zentral oder dezentral beispielsweise durch jeden Web-Site-Autor selbst organisiert werden.

Wenn der geographische Bezug einer Web-Site sich ändert, müssen auch die Daten der Seiten geändert und in der Suchmaschine aktualisiert werden. Abhängig davon, wie die Geoinformation gespeichert wurde, muss der Prozess der Aktualisierung für jede Änderung manuell durchgeführt werden. Er kann jedoch auch automatisch ablaufen und muss nur gegebenenfalls manuell angestoßen werden.

Durch Genauigkeit, Vollständigkeit und Eindeutigkeit der Repräsentation des geographischen Bezugs wird die Güte der Ergebnisse beeinflusst. Eine Suchmaschine mit einer hohen Fehlerquote verliert an Nutzen und Akzeptanz bei den Anwendern.

Der hier betrachtete Einsatzbereich der geographischen Suche kann entweder das Web bzw. das Internet oder das Intranet sein. Je nach Anwendungsbereich bestehen unterschiedliche Anforderungen an die Skalierbarkeit des Konzeptes.

Neben den Methoden, den geographischen Bezug zu repräsentieren und in der Suchmaschine zu speichern, können verschiedene Möglichkeiten der Präsentation unterschieden werden (s. Abb. 3.2) Die Ergebnisse müssen nach individuellen Präferenzen konfigurierbar dargestellt und sortierbar sein.

Die Präsentation wird durch die Art des Endgerätes stark beeinflusst. Mobile Endgeräte sind typischerweise klein und verfügen nicht über die gleichen graphischen Möglichkeiten wie ein Computerbildschirm. Auch die begrenzte Bandbreite muss bei der Auswahl der Darstellung berücksichtigt werden.

Um eine geographische Suchanfrage zu stellen, muss der räumliche Suchbereich festgelegt werden. Die Ortsangabe kann über die textuelle Eingabe einer Adresse geschehen, aber auch graphisch auf einer Karte ausgewählt werden. Wird ein mobiles Endgerät verwendet, kann die eigene Position durch GPS lokalisiert und als Ausgangspunkt der Suche verwendet werden.

Endgerät	mobil		nicht-mobil
Festlegung des Suchbereichs	textuell	graphisch	automatisch
Präsentation I	kartenverwandte Darstellung	fotografische Darstellung	textuelle Darstellung
Präsentation II	interaktiv		passiv
Ordnungskriterium der Ergebnisse	geographisch	inhaltlich	zeitlich

Abb. 3.2: Zweiter Teil des Morphologischen Kastens

Je nach persönlicher Präferenz kann zur Präsentation der Ergebnisse eine kartenverwandte, fotografische oder textbasierte Darstellung gewählt werden. Sollen die Karten nach individuellen Bedürfnissen konfigurierbar und die Ergebnisse sortierbar sein, so spricht man von einer interaktiven Präsentation. Als mögliche Ordnungskriterien der Ergebnisse kommen geographische Aspekte wie die Position auf einer Karte oder die Entfernung vom eigenen Standpunkt, aber auch inhaltliche Kriterien in Frage. Unter

zeitlichen Ordnungskriterien werden Veröffentlichungsdatum oder ein Zeitraum verstanden.

### **3.1 Endkunden-orientierte Anwendung**

Eine geographische Suchmaschine eignet sich besonders für Anwendungsfälle, in denen die räumliche Lage und Entfernungen eine wesentliche Rolle spielen.

#### **"Tourismus"**

Die Orientierung in fremden Städten wird durch die unbekannte Gegend erschwert. Auf dem Spaziergang durch fremde Straßen und Gassen entdeckt man interessante Sehenswürdigkeiten, über die man gerne mehr erfahren möchte. Häufig sind diese Informationen im Internet vorhanden und müssen nur in geeigneter Weise auffindbar gemacht werden. Um eine Suchanfrage an eine herkömmliche textbasierte Suchmaschine stellen zu können, werden Stichworte wie der Name oder markante Hintergrundinformationen, die als geeignete Suchbegriffe dienen könnten, benötigt. Da diese meist nicht bekannt sind oder als alleinige Suchkriterien nicht ausreichen, kann durch die neue Art der Suche auch die geographische Position der Sehenswürdigkeit als Suchkriterium eingesetzt werden.

In diesem Anwendungsszenario werden Informationen über die Kaiser-Wilhelm-Gedächtniskirche in Berlin gesucht, deren Lage in der Nähe des Tiergartens bekannt ist. Da es viele Kirchen in Berlin gibt, scheiden die Suchwörter "Kirche" und "Berlin" als alleinige Suchbegriffe aus.

Die Web-Site der geographischen Suchmaschine wird aufgerufen. Von der Deutschlandkarte aus wird zunächst nach Berlin navigiert (s. Abb. 3.3), um dort den Kartenausschnitt Tiergarten / Kurfürstendamm auszuwählen (s. Abb. 3.4).

Wenn der Suchraum festgelegt ist, kann der Suchmaschine ein Suchobjekt angegeben werden. Dies könnte in diesem Fall der Begriff "Kirche" sein, da der genaue Name der Kaiser-Wilhelm-Gedächtniskirche nicht bekannt ist. Der Suchvorgang kann gestartet werden.

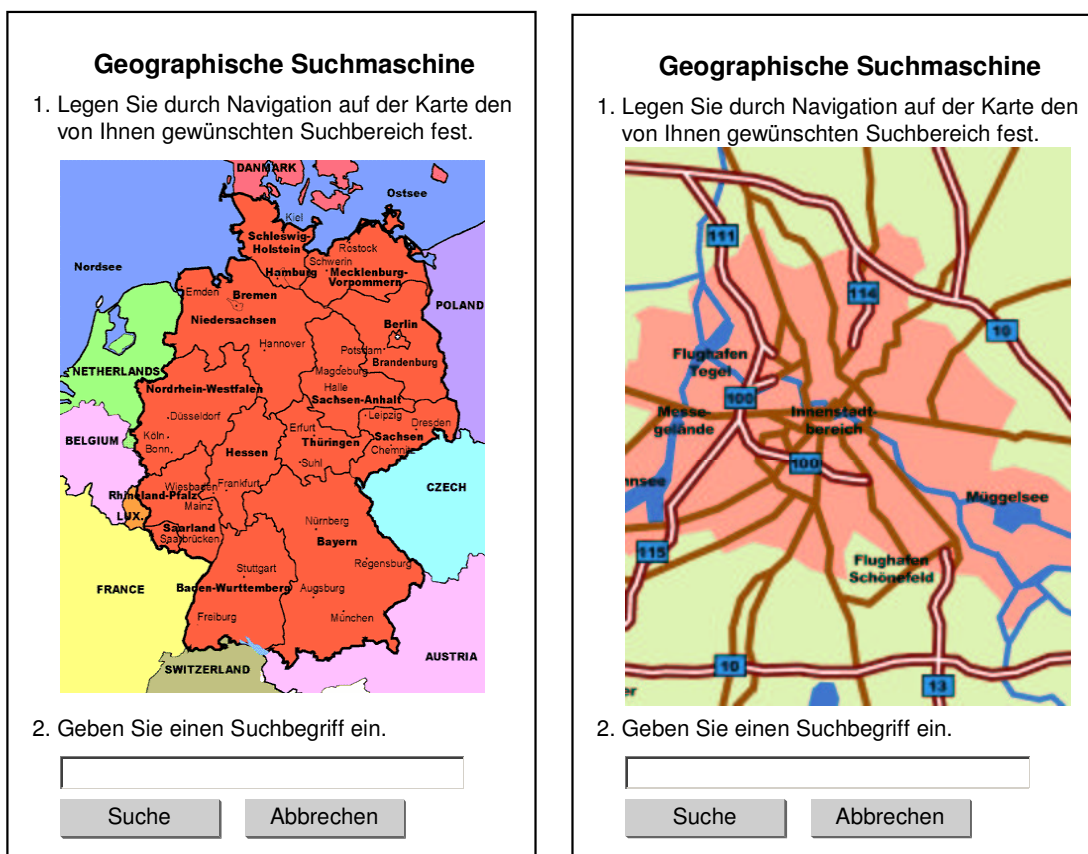


Abb. 3.3: Suchablauf einer Endkunden-orientierten Anwendung [1]

Die Informationsakquisitionskomponente der Suchmaschine durchsucht kontinuierlich das Web nach Seiten mit geographischem Bezug und nimmt diese in ihren Datenbestand auf. Die geographische Referenz wird vorher vom Web-Site-Autor mit der Seite verknüpft.

Auf dieser Datenbasis führt die Suchmaschine Anfragen aus und gibt als Ergebnis diejenigen Seiten aus, deren geographischer Bezug innerhalb des Kartenausschnitts liegt und die den Suchbegriff enthalten. Durch das geographische Kriterium werden also sämtliche Web-Sites über Kirchen, die nicht auf einen Bereich innerhalb des Kartenausschnitts verweisen, ausgeschlossen.

Die Suchergebnisse können nach unterschiedlichen Aspekten sortiert werden. Ist die Lage das Ordnungskriterium, werden die auf das Zentrum des Ausschnitts verweisen den Web-Sites an den Anfang sortiert. Aber auch die Größe der Schnittfläche des Kartenausschnitts mit dem Referenzbereich der Web-Seite kann als Kriterium herangezogen werden.

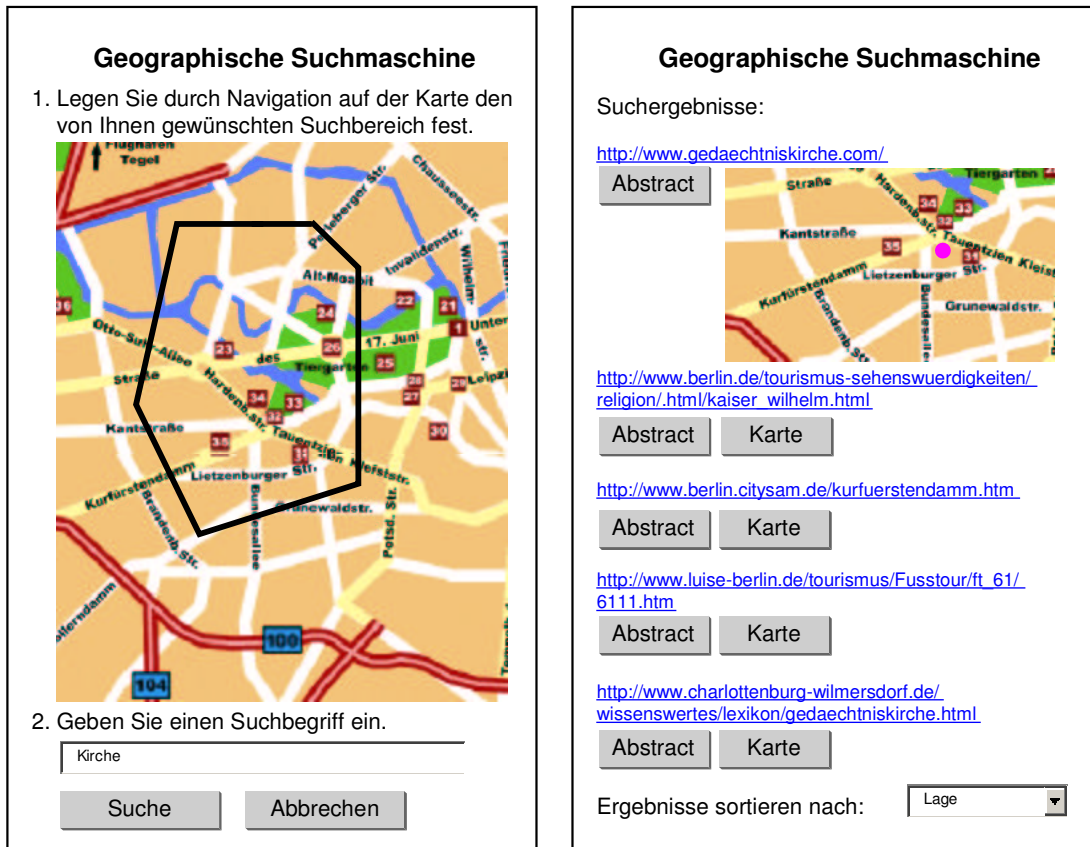


Abb. 3.4: Suchablauf einer Endkunden-orientierten Anwendung [2]

Für jeden Treffer besteht die Möglichkeit, den geographischen Bezug in einem gesonderten Kartenausschnitt zu visualisieren. Zur Vorauswahl der Suchergebnisse kann außerdem eine inhaltliche Zusammenfassung genutzt werden. So surft der Nutzer bequem über den Link auf die Seiten, auf denen die gesuchte Information zu finden ist.

### Analyse

In diesem Szenario wird die geographische Information auf den Web-Sites von dem Autor ohne Eingriff von außen seitens der Suchmaschine kodiert, daher ist dies eine direkte Informationserhebung.<sup>14</sup> Die Eingabe und die Pflege wird also ausschließlich dezentral durch die Autoren vorgenommen, die bei Änderungen der geographischen Referenz die Daten auf ihrer Web-Site aktualisieren. Durch die Suchmaschine werden die neuen Daten beim nächsten Suchdurchlauf automatisch in der Datenbasis aktualisiert. Durch die explizite Speicherung der räumlichen Daten kann ein für die Suchmaschine interpretierbares Format festgelegt werden, so dass fehlerhafte Suchergebnisse aufgrund von falsch interpretierten Daten auf der Web-Site nicht vorkommen. Nur

<sup>14</sup> Zu den Möglichkeiten der Realisierung der direkten, indirekten und automatischen Informationserhebung wird in Kap.4.1 bis Kap.4.3 detailliert eingegangen.

durch falsche, ungenaue bzw. zu große Raumangaben, um beispielsweise die Anzahl der Hits auf der eigenen Homepage zu erhöhen, können Fehler entstehen.

Dieses Konzept der direkten Informationserhebung ermöglicht durch seine dezentrale Datenpflege eine flexible Skalierung bezüglich des Umfangs der erfassten Daten und einen niedrigen Aufwand auf Verwalterseite. Es eignet sich daher auch für den Einsatz in dem ständig wachsenden Internet.

Der Zugriff auf die Suchmaschine erfolgt je nach Situation über ein mobiles oder nicht-mobiles Endgerät. Um beispielsweise einen Ausflug vor- oder nachzubereiten, wird die Suchmaschine von einem nicht-mobilen Endgerät verwendet. Ihr ganzes Potenzial wird jedoch erst bei mobiler Nutzung deutlich. Auf einer Reise können spontan Erkundigungen zu Gebäuden oder Sehenswürdigkeiten in der eigenen Umgebung eingezogen werden.

Informationserhebung	direkt	indirekt	automatisch
Datenpflege	zentral		dezentral
Aktualisierung des Geo-Bezugs	manuell	Aktivierung manuell, Durchführung automatisch	automatisch
Fehlerquote der Suchmaschine	gering		hoch
Anwendungsbereich	Internet		Intranet
Skalierbarkeit	hoch		beschränkt
Endgerät	mobil		nicht-mobil
Festlegung des Suchbereichs	textuell	graphisch	automatisch
Präsentation I	kartenverwandte Darstellung	fotografische Darstellung	textuelle Darstellung
Präsentation II	interaktiv		passiv
Ordnungskriterium der Ergebnisse	geographisch	inhaltlich	zeitlich

Abb. 3.5: Analyse der Endkunden-orientierten Anwendung

Im obigen Szenario wird die Abfrage kombiniert aus einer geographischen Eingrenzung des Suchbereichs und einem textuellen Suchwort durchgeführt, wobei auch eine rein geographische Suche denkbar ist. Zur Präsentation der Suchergebnisse wird eine textuelle Darstellung genutzt, die durch eine Karte, die zuvor auch die Suchanfrage unterstützt, ergänzt werden kann. Zur Sortierung werden geographische Kriterien angeboten.

### 3.2 Business-orientierte Anwendung

In diesem Szenario wird die geographische Referenz als eine alternative Ordnungsstruktur für den firmeninternen Datenbestand dargestellt.

#### "Störung und Wartung"

In einem großen Chemiewerk kommt es zu einem Störfall, zu dem ein Techniker zur Reparatur gerufen wird. Da die Angaben der Störungsmeldung zu ungenau sind, um den ganzen Ablauf im Voraus zu planen, fährt der Techniker erst zum Ort des Geschehens, um sich dort genauer zu informieren.

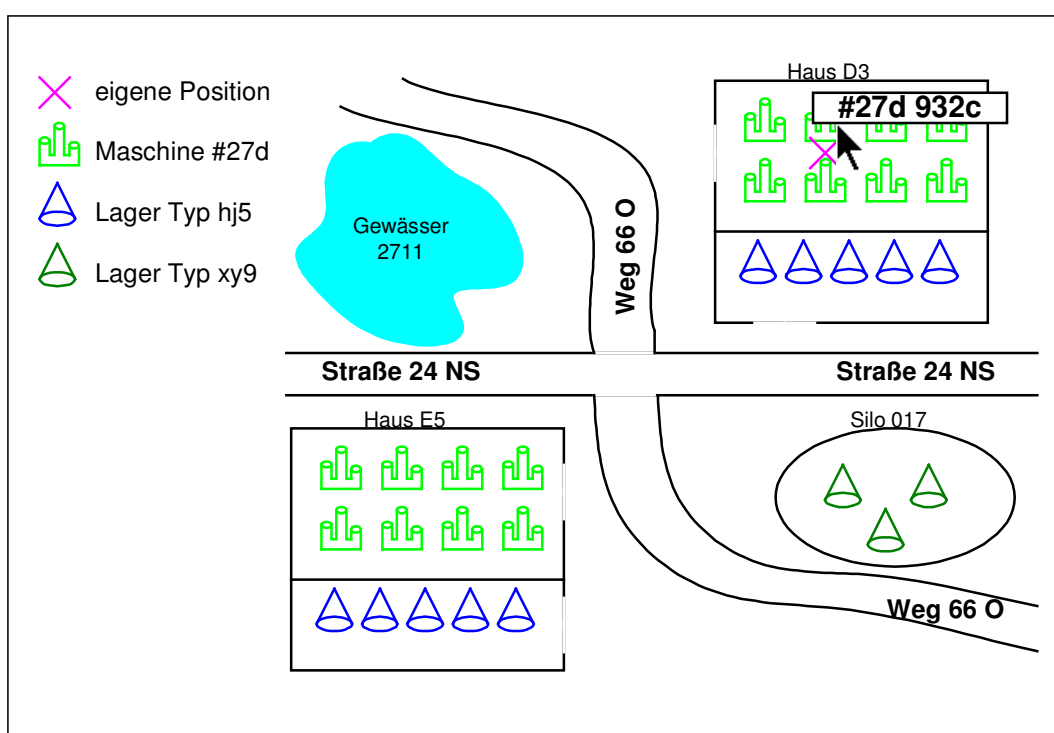


Abb. 3.6: Ergebnisdarstellung einer Business-orientierten Anwendung

Der Techniker steht vor der defekten Maschine. Für Wartung und Reparatur existieren diverse Vorschriften zu Umfang und Vorgehensweise, die im Dokumentenmanagementsystem der Firma hinterlegt sind. Mittels GPS wird seine Position bestimmt und der entsprechende Kartenausschnitt vom Firmengelände in dem vom Techniker angegebenen Maßstab angezeigt (s. Abb. 3.6).

Die Karte kann manuell durch den Techniker konfiguriert werden, so dass die benötigten Informationen wie Wege und Maschinen, aber auch Strom-, Wasserleitungen etc. angezeigt werden. Durch Anklicken des Icon der defekten Maschine wird automatisch eine Suchanfrage mit der Identifikationsnummer des Maschinentyps als Stichwort und den Positionskordinaten als geographisches Suchkriterium generiert.

Jedes der Dokumente im Dokumentenmanagementsystem der Firma ist mit der Information versehen, auf welchen Bereich des Firmengeländes sich der Inhalt bezieht. Ein Werkzeug oder ein Add-in unterstützt den Autor eines Dokumentes in der Auswahl der Bezugspunkte auf der Karte des Firmengeländes, um seine Datei entsprechend ihrer Inhalte richtig zu verlinken. Mit der Suchanfrage werden die Angaben zur räumlichen Lage der Maschine übermittelt und alle Dokumente, deren Inhalt sich auf die geographische Position der Maschine bezieht, als Ergebnis zurückgeliefert.

Wenn die Trefferliste zu umfangreich ist, spezifiziert der Techniker die Anfrage durch einen weiteren Suchbegriff wie z.B. Entsorgungsrichtlinien, Umweltverordnung, Schutzkleidung, Reparaturanleitung und lässt die Ergebnisse entsprechend inhaltlich sortieren.

### **Analyse**

Für die Festlegung des geographischen Bezugs in diesem Szenario ist der Autor eines jeden Dokuments zuständig. Folglich findet hier wie bei der Endkunden-orientierten Anwendung eine direkte Informationserhebung statt, deren Daten dezentral von den Mitarbeitern selbst eingepflegt werden. Falls der geographische Bezug sich ändert, kann der Verfasser des Textes durch das Tool direkt die Daten in dem Dokumentenmanagementsystem ändern.

Als Intranetlösung für eine Firma existieren zur Informationscodierung feste Vorschriften, deren Einhaltung durch ein Werkzeug gesichert wird, so dass abgesehen von falscher Eingabe der Daten mit fehlerlosen Ergebnissen zu rechnen ist. Dieses Konzept ist durch die dezentrale Erhebung der Daten auch skalierbar und für das Internet geeignet.

Der Zugriff auf die Dokumente erfolgt hier mittels eines mobilen Endgerätes, welches über den Global Positioning Service (GPS) die eigene Position ortet und als Ausgangspunkt für die Suche verwendet. So wird der Suchbereich automatisch durch die Position und einen gewünschten Radius festgelegt und die Navigation auf einer Karte entfällt.

Informationserhebung	direkt	indirekt	automatisch
Datenpflege	zentral		dezentral
Aktualisierung des Geo-Bezugs	manuell	Aktivierung manuell, Durchführung automatisch	automatisch
Fehlerquote der Suchmaschine	gering		hoch
Anwendungsbereich	Internet		Intranet
Skalierbarkeit	hoch		beschränkt
Endgerät	mobil		nicht-mobil
Festlegung des Suchbereichs	textuell	graphisch	automatisch
Präsentation I	kartenverwandte Darstellung	fotografische Darstellung	textuelle Darstellung
Präsentation II	interaktiv		passiv
Ordnungskriterium der Ergebnisse	geographisch	inhaltlich	zeitlich

Abb. 3.7: Analyse der Business-orientierten Anwendung

Die Präsentation der Ergebnisse erfolgt über eine Karte, auf der die Umgebung durch Symbole dargestellt wird. Diese Karte ist je nach Anwendungskontext interaktiv konfigurierbar - Bodenbeschaffenheit, Wasserläufe oder Läger mit leicht entzündlicher Ware können angezeigt werden. Die Suchergebnisse können nach inhaltlichen Kriterien sortiert werden.

### 3.3 Administrations-orientierte Anwendung

Der geographische Bezug ordnet Objekten eine bestimmte Position im Raum zu. Hier muss es sich nicht immer um reale oder sichtbare Gegenstände handeln: auch Pläne von einem zukünftigen Gebäude oder unterirdisch verlegte Leitungen haben einen Bezug, der durch eine Suchmaschine visualisiert wird.

#### "Bauamt"

Ein Architekt möchte ein Haus bauen. Damit der Wert des Baugrundstücks richtig beurteilt und die Pläne für das Haus den Gegebenheiten angepasst werden können, interessieren ihn sämtliche Informationen der Stadt über Bauordnungen, Richtlinien, Genehmigungen und Mietpreise sowie die Topologie und Bebauung des Umfeldes. Wesent-

lich für die Planung eines Hauses ist auch die vorhandene Infrastruktur wie Wasser- und Gasanschluss und möglicherweise innerhalb des Grundstücks verlegte Leitungen.

Angestellte der Stadt registrieren alle vorhandenen und geplanten Bauprojekte und Infrastrukturen in einem Geoinformationssystem. In dieser monolithischen Anwendung werden sämtliche lokale Daten abgelegt und über eine Web-Schnittstelle vom Internet aus zugänglich gemacht. Informationen wie Ort des Vorhabens, Zeitdauer und Ablauf, detaillierte Angaben über Größe, Höhe und Art der Bauten werden eingetragen. Um die Konsistenz und die Richtigkeit zu sichern, überprüft ein Redaktionsteam die Daten vor der Veröffentlichung.

### Geographische Suchmaske

1. Geben Sie die Adresse ein, in deren Umfeld Sie die Bauvorhaben der Stadt interessieren.

Straße\*

Hausnummer\*

Viertel

Ort\*

\* optional

2. Bestimmen Sie die Größe des Umfeldes:

Umkreis:

oder bestimmen Sie den Bereich graphisch:

Abb. 3.8: Suchanfrage einer Administrations-orientierten Anwendung

Will ein Architekt beispielsweise Informationen zu einem Bauvorhaben sammeln, gibt er die Adresse des gewünschten Baugrundstücks an (s. Abb. 3.9) und legt den Radius des Umkreises fest.

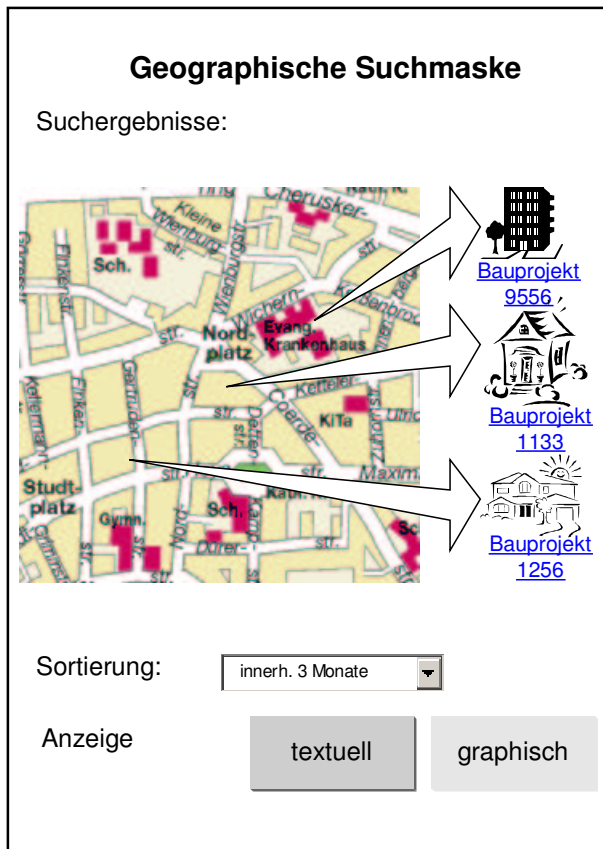


Abb. 3.9: Ergebnisdarstellung einer Administrations-orientierten Anwendung

Die Adresse wird nun mit Hilfe eines Geo-Coders in Koordinatenangaben umgerechnet. Sämtliche Baupläne werden durchsucht und die relevanten Pläne in dem vom Architekten festgelegten Kartenausschnitt symbolisch visualisiert.

Alternativ kann auch eine textuelle Darstellung gewählt werden. Zur Übersicht können die Ergebnisse nach verschiedenen Kriterien wie Aktualität, Dauer oder Thematik sortiert werden.

Über Links kann der Architekt direkt das entsprechende Dokument im Datenbestand des Bauamts aufrufen.

Ein besonderer Aspekt dieses Szenarios besteht darin, dass durch den räumlichen Bezug etwas visualisiert werden kann, das bisher noch nicht existiert. Auch Bus- und Bahnlinien, sowie geplante Autobahnen und die Parkplatzsituation können für die Auswahl der Arbeitsstätte oder Investitionen in einen Standort von Relevanz sein.

### Analyse

Um die Bauvorhaben der Stadt und sämtliche Infrastruktur mit Geoinformationen zu versehen, wird ein speziell hierfür entwickeltes Geoinformationssystem verwendet, in

dem in einer vorgegebenen Maske die Daten eines Bauprojektes wie Ort, Dauer und Beginn des Vorhabens eingegeben werden. Diese Daten sind über die Web-Schnittstelle der Anwendung vom World Wide Web aus zugänglich.

Um eine hohe Qualität der Daten sicherzustellen, wird jeder Eintrag von einem Redaktionsteam geprüft. Da der Autor nicht selbst die Daten in das Geoinformationssystem einträgt, sondern Dritte an diesem Prozess beteiligt sind, wird dies als indirekte Informationserhebung bezeichnet. Die zentrale Datenpflege durch Angestellte der Stadt erfordert bei einem großen Datenbestand einen steigenden personellen Aufwand und ist somit nur auf eine beschränkte Datenmenge anwendbar. Für das World Wide Web ist dieser Ansatz ungeeignet. Im Falle von Änderungen muss die Aktualisierung zentral und manuell vorgenommen werden.

Informationserhebung	direkt	indirekt	automatisch
Datenpflege	zentral		dezentral
Aktualisierung des Geo-Bezugs	manuell	Aktivierung manuell, Durchführung automatisch	automatisch
Fehlerquote der Suchmaschine	gering		hoch
Anwendungsbereich	Internet		Intranet
Skalierbarkeit	hoch		beschränkt
Endgerät	mobil		nicht-mobil
Festlegung des Suchbereichs	textuell	graphisch	automatisch
Präsentation I	kartenverwandte Darstellung	fotografische Darstellung	textuelle Darstellung
Präsentation II	interaktiv		passiv
Ordnungskriterium der Ergebnisse	geographisch	inhaltlich	zeitlich

Abb. 3.10: Analyse der Administrations-orientierten Anwendung

In diesem Szenario wurde die Möglichkeit einer rein textuellen Anfrage über eine Adresse vorgestellt, aber auch eine graphische Unterstützung ist denkbar. Die Angabe der Adresse kann in geographische Koordinaten umgerechnet werden. Polygone, die auf einer Karte durch eine Maus festgelegt werden, können hier durch Angabe des Umkreisradius als Kreis dargestellt werden. Da die Suchergebnisse auf einer interaktiv konfigurierbaren, teilweise sehr detaillierten Karte dargestellt werden, ist die Verwendung eines großen Bildschirms von Vorteil. Zur besseren Lesbarkeit der Grafik ist also ein

nicht-mobiles Endgerät geeignet, grundsätzlich ist aber auch der mobile Zugriff möglich. Hierfür dürfte sich beispielsweise auf Baustellen ein Bedarf ergeben.

## 4 Zusammenstellung und Diskussion konzeptioneller Ansätze

Eine geographische Suchmaschine muss verschiedene Funktionalitäten haben: Sie muss, um Suchanfragen bearbeiten zu können, die im Web hinterlegten geographischen Informationen zunächst extrahieren. Diese Informationen werden in der Informationsbasis abgespeichert, um später Abfragen effizient verarbeiten zu können. Die Ergebnisse zu einer Anfrage muss die Suchmaschine in geeigneter Form dem Nutzer darstellen.

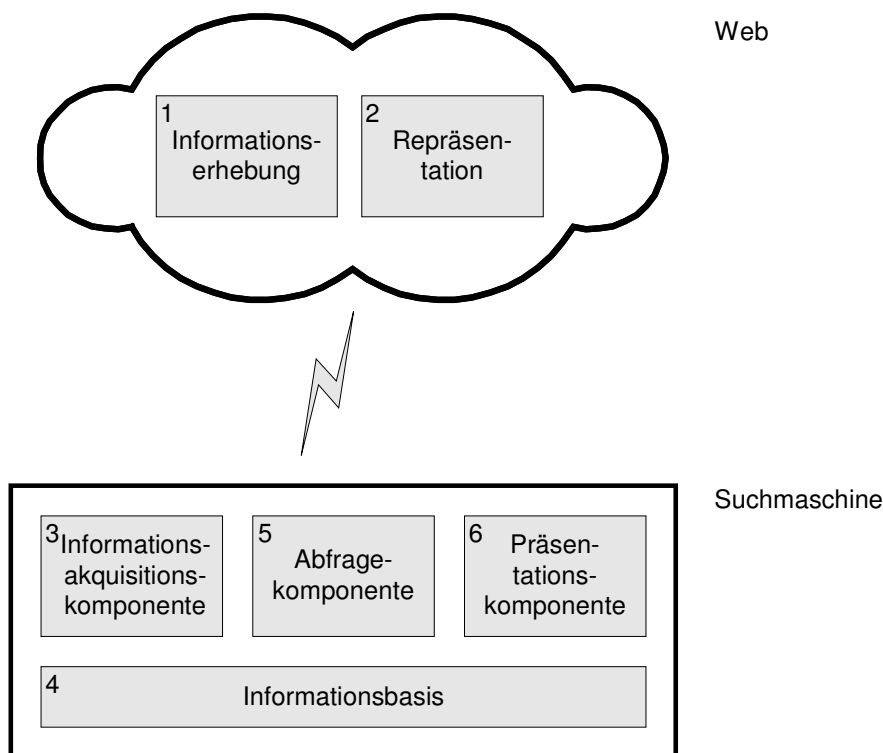


Abb. 4.1: Struktur einer geographischen Suche

In Abb. 4.1 wird eine Struktur vorgeschlagen, die eine geographische Suche ermöglicht. Die Wolke im oberen Teil des Bildes repräsentiert das Web und der Kasten stellt eine hoch aggregierte Architektur der geographischen Suchmaschine dar.

Zunächst muss der geographische Bezug einer Web-Site festgestellt werden. Als Informationsquelle dient der Web-Site-Autor oder vorhandene Informationen, aus denen auf den geographischen Bezug der Web-Site geschlossen werden kann. Im folgenden Schritt muss eine geeignete Repräsentation dieser Informationen innerhalb einer Homepage gefunden werden.

Sind die Geoinformationen im Web vorhanden, müssen suchmaschinenspezifische Aufgaben gelöst werden. Die Architektur einer geographischen Suchmaschine gliedert sich in vier Komponenten, die die verschiedenen Funktionalitäten abdecken. Dieser Archi-

tekturvorschlag orientiert sich an dem Aufbau eines Expertensystems.<sup>15</sup> Die Informationsakquisitionskomponente extrahiert die im Web gespeicherten geographischen Informationen, um sie strukturiert in der Informationsbasis abzulegen. Bei der Suchanfrage eines Benutzers filtert die Abfragekomponente die relevanten Web-Sites als Suchergebnisse heraus, welche durch die Präsentationskomponente dem Nutzer in einer adäquaten Weise dargestellt werden.

Im Folgenden werden die Teilschritte zur Realisierung einer geographischen Suche vertieft.

#### **4.1 Informationserhebung im Web**

Unter der Informationserhebung versteht man den Prozess der "Extraktion" von Information aus Informationsquellen sowie die Überführung in eine formale Sprache. Die einzelnen Teilschritte der Informationserhebung für eine geographische Suche werden in drei getrennten Kapiteln behandelt. Im vorliegenden Kapitel wird die Extraktion und Explikation des geographischen Bezugs einer Web-Site beschrieben. Es wird untersucht, welche Personen oder Maschinen den Bezug feststellen können und welche Methoden hierzu angewendet werden. In Kapitel 4.2 dagegen wird die Überführung der Geoinformation in eine formale Sprache dargestellt. Wie diese Information wiederum einer Suchmaschine zugänglich gemacht wird, ist Thema des Kapitels 4.3.

Bei Expertensystemen werden drei Grundmodelle der Wissensakquisition unterschieden.<sup>16</sup> In Anlehnung an diese Modelle wird auch die Informationserhebung der geographischen Suche in die direkte, indirekte und automatische Erhebung unterteilt.

Die *direkte Informationserhebung* liegt vor, wenn die Geoinformation aus der Informationsquelle gelesen wird und ohne Zwischenschalten eines Dritten in der Informationsbasis der Suchmaschine abgelegt wird.

Als Informationsquelle dient bei der geographischen Informationserhebung der Autor einer Web-Site. Voraussetzung für diese Form der Informationserhebung ist, dass der Autor ein ausreichendes Verständnis für die formale Darstellung der Daten hat, so dass die Informationsakquisitionskomponente der Suchmaschine diese lesen kann. Verschiedene Realisierungsmöglichkeiten der Akquisitionskomponente werden in Kap. 4.3 detailliert beschrieben.<sup>17</sup>

---

<sup>15</sup> Vgl. Kurbel, K. (1992), S. 28.

<sup>16</sup> Vgl. Schneider, B.; (2001), S. 7.

<sup>17</sup> Vgl. Kurbel, K. (1992), S. 70.

Wenn die Informationserhebung über einen Mittelsmann erfolgt, spricht man von *indirekter Informationserhebung*. Dieser Vermittler expliziert die Geoinformationen und bringt sie in ein geeignetes Repräsentationsschema. Diese manuelle Vorgehensweise ist jedoch mit erheblichem Aufwand verbunden.

Bei der geographischen Suchmaschine gibt es keine direkte Parallele zu dem Mittelsmann der Expertensysteme. Trotzdem können an dem Informationserhebungsprozess auch Dritte beteiligt sein: z.B. kann die Erhebung durch einen Lektor geschehen, der sich über die Web-Site informiert. Außerdem können die von einem Autor oder Lektor erhobenen Informationen durch ein Redaktionsteam, das die Angaben überprüft und modifiziert, bearbeitet werden.<sup>18</sup>

*Automatische Informationserhebung* bedeutet, dass die Suchmaschine ihre Informationen ohne menschliche Hilfe automatisch erwirbt.

Die Daten werden von einem Programm analysiert, die Information extrahiert und in geeigneter Form in der Informationsbasis gespeichert. Auf den meisten Web-Sites existieren bereits Daten, die Informationen über den geographischen Bezug des Inhalts enthalten.<sup>19</sup>

Die direkte und die indirekte Informationserhebung werden hier als manuelle Informationserhebung zusammengefasst und der computergestützten, automatischen Informationserhebung gegenübergestellt.

### **Anforderungen an die Informationserhebung im Web**

Damit die Informationserhebung auf der Ebene des Webs eine weite Verbreitung findet, muss sie verschiedenen Anforderungen gerecht werden:

- Alle befragten Informationsquellen müssen ein gleiches oder zumindest ähnliches Verständnis vom geographischen Bezug haben und davon, was sie inhaltlich codieren wollen. Die verschiedensten Angaben wie Standort einer Firma, Inhalt einer Web-Site, Einzugsgebiet der Kunden oder Auswirkungsgebiet von Projekten können als geographischer Bezug einer Web-Site benutzt werden. Bei der Informationserhebung muss also den unterschiedlichen Interpretationen Rechnung getragen werden.
- Das Konzept der Informationserhebung muss skalierbar sein, damit es für weite Teile des Internet anwendbar ist. Damit die Masse von Geoinformationen im Web an-

---

<sup>18</sup> Vgl. Kurbel, K. (1992), S. 68.

<sup>19</sup> Vgl. Kurbel, K. (1992), S. 70-71.

nähernd vollständig analysiert werden kann, wird ein leistungsfähiges System benötigt.

Durch Wahlmöglichkeit zwischen verschiedenen Verfahren kann der Anwender die von ihm favorisierte Lösung nutzen.

### **Konzepte für die Informationserhebung im Web**

Zur Informationserhebung müssen zunächst mögliche Informationsquellen identifiziert werden, um dann ein adäquates Anfrageverfahren zu finden.

#### *Manuelle Informationserhebung*

Bei der manuellen Informationserhebung dient der Web-Site-Autor, der sowohl den Inhalt der Site kennt als auch Hintergrundinformationen über Entstehung, Motivation und Ziele der Web-Site hat, als Informationsquelle. Bei dieser dezentralen Informationserhebung entsteht durch die große Zahl an Web-Site-Autoren unvermeidlich ein individuelles Verständnis vom geographischen Bezug einer Web-Site. Da die geographischen Daten eigens für diesen Zweck erhoben werden, kann Inhalt und Format der Daten bestimmt werden, um so eine möglichst einheitliche Codierung vorzugeben.

Statt Web-Site-Autoren können auch Dritte, z.B. Lektoren, die benötigten Informationen liefern. Sie informieren sich auf der Homepage oder über die Kontaktadresse auf der Web-Site bei dem Ansprechpartner über die dargestellten Inhalte und legen den geographischen Bezug fest. Ein Team von Lektoren spricht sich vorher ab, was unter einem geographischen Bezug zu verstehen ist und legt eine einheitliche Codierung dieses Bezugs fest.

#### *Automatische Informationserhebung*

Die automatische Informationserhebung nutzt Informationen aus anderen Bereichen, um von dort aus automatisch auf den geographischen Bezug schließen zu können. An diesem Prozess der Informationserhebung ist kein Mensch beteiligt. Beispielsweise kann die Standortwahl eines Web-Servers Aufschluss über einen geographischen Bezug liefern, indem die IP-Adresse oder das Landeskürzel der URL untersucht werden. Auch Kontaktadressen oder Anfahrtsbeschreibungen sind Geoinformationen, die auf vielen Web-Sites bereits vorhanden sind. Sie geben implizit Auskunft über einen möglichen geographischen Bezug.

### **Vergleich der Informationserhebungskonzepte**

Wenn eine manuelle Informationserhebung durch den *Autor einer Web-Site* erfolgt, kann dieses dezentrale Verfahren problemlos auf große Datenmengen wie sie im World

Wide Web vorhanden sind, skaliert werden. Entscheidend für den Erfolg dieses Verfahrens ist die Art der Codierung der Geoinformation. Sind die anzugebenden Daten genau und eindeutig beschrieben, kann eine hohe Qualität der Daten erreicht werden, die somit sinnvoll von der Suchmaschine genutzt werden können. Da eine Suchmaschine erst durch die Menge der in ihr verzeichneten Web-Ressourcen attraktiv wird, ist die Akzeptanz dieses Verfahrens durch die Web-Site-Autoren von elementarer Bedeutung.

Das einheitliche Verständnis vom geographischen Bezug einer Web-Site ist durch die kleine Gruppe von Lektoren sehr gut zu lösen. Das Lektorenteam kann allerdings durch seine beschränkte Kapazität nur einen sehr geringen Teil des Webs bearbeiten. Da hier nach einem Konzept gesucht wird, das auf das Web angewendet werden kann, spielt der Lektor als Informationsquelle keine Rolle und wird im Folgenden nicht weiter berücksichtigt.

Die *automatische Informationserhebung* durch implizite Informationen hat den Vorteil, dass Informationen, die bereits verfügbar sind, genutzt werden können. Jedoch stimmen die Rückschlüsse, die z.B. aus dem Standort des Servers gezogen werden, nicht unbedingt mit dem tatsächlichen geographischen Bezug des Inhalt auf dem Web-Server, überein. Diese Problematik wird im Detail in Kap. 4.2 beschrieben.

## **4.2 Repräsentation im Web**

Nachdem der geographische Bezug festgelegt worden ist, muss die Information adäquat im Web repräsentiert werden, um eine effiziente Informationsakquisition der Suchmaschine zu realisieren. Die verschiedenen Möglichkeiten der Codierung werden im Folgenden vorgestellt.

### **Anforderungen an die Repräsentation im Web**

Um eine breite Akzeptanz zu erreichen, muss die Codierung einigen Anforderungen genügen:

- Die wichtigste Eigenschaft einer Codierung ist ihre Eindeutigkeit: es müssen möglichst alle unterschiedlichen Interpretationsmöglichkeiten eines geographischen Bezugs ausgeräumt werden.
- Die Repräsentationsform der Geoinformationen muss benutzerfreundlich sein - die Akzeptanz kann beispielsweise durch ein übersichtliches Werkzeug gefördert werden. Das Konzept muss schnell durchschaubar und einfach durchzuführen sein. Eine

lange Einarbeitungszeit und zu umfangreiche Daten würden zu sinkender Motivation und geringer Akzeptanz führen.

## Konzepte für die Repräsentation im Web

Die unterschiedlichen Konzepte der manuellen und der automatischen Informationserhebung werden in Abb. 4.2 dargestellt. Sie sind nach ihrem Strukturierungsgrad sortiert.

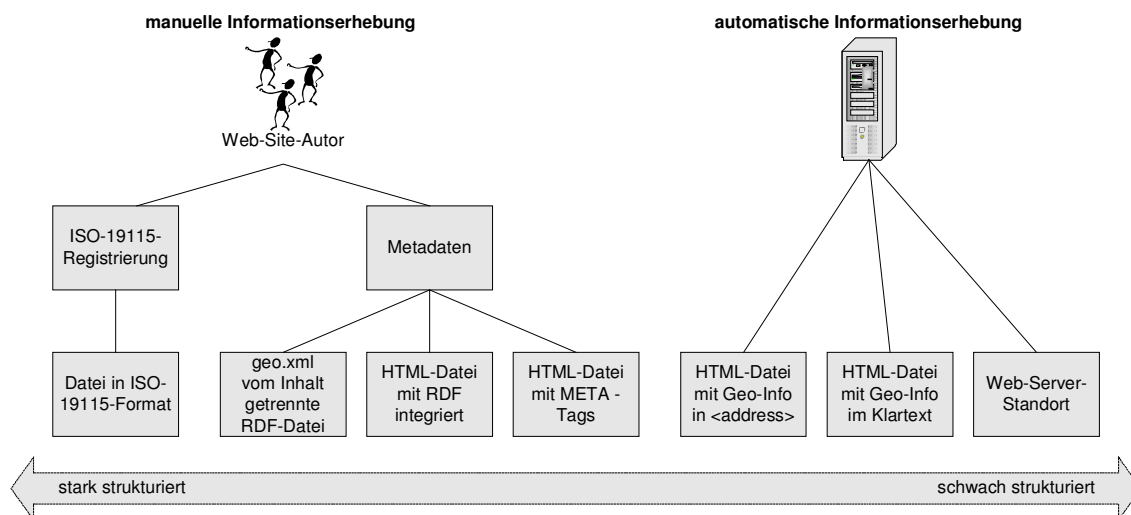


Abb. 4.2: Informationserhebung und Repräsentation im Internet

### *ISO-19115-Registrierungsdienst*

Der auf dem ISO-19115-Standard basierende Registrierungsdienst kommt bei manueller Informationserhebung zum Einsatz. Der Web-Site-Autor kann seine Web-Site durch ein öffentlich zugängliches Formular bei einer Suchmaschine registrieren. Die Inhalte dieses Formulars basieren auf dem ISO-19115-Standard "Geographic information - Metadata". Es besteht aus optionalen und verpflichtenden Feldern, in denen Angaben zur Homepage und ihrem geographischen Bezug gemacht werden.

Die über einen zentralen Dienst erhobenen Daten liegen im standardisierten, maschinenlesbaren ISO-19115-Format vor und können so in einer Datenbasis abgelegt werden.

### *Metadaten*

Alternativ zu einem zentralen Registrierungsdienst kann der geographische Bezug einer Web-Site auch durch Metadaten gespeichert werden. Format und Bedeutung der Metadaten müssen vorher für alle möglichen Raumbezugssysteme wie beispielsweise Koordinaten und Adressen festgelegt werden.

Zur Implementierung von Metadaten eignen sich das META-Tag von HTML oder das Resource Description Framework (RDF). In RDF codierte geographische Daten können entweder in einer eigenen XML-Datei, die einen Verweis auf das beschriebene Doku-

ment beinhaltet, gespeichert werden oder aber in die zu beschreibende HTML-Datei integriert werden. META-Tags werden in den HEAD-Teil eines HTML-Dokuments geschrieben.

#### *HTML-Datei mit Geoinformationen*

Auf den meisten Web-Seiten existieren bereits heute implizite geographische Angaben, aus denen häufig sinnvolle Rückschlüsse auf den geographischen Bezug einer Web-Site gezogen werden können. Solche Angaben sind z.B. eine Lage- oder Anfahrtsbeschreibung, eine Kontaktadresse, die entweder in dem von HTML dafür vorgesehenen <address> Tag oder im Klartext der Datei steht, oder es werden einfach Ortsangaben innerhalb des Textes zur Analyse herangezogen. Verglichen mit der fest vorgegebenen Struktur des ISO-19115-Registrierungsdienstes oder den Metadaten sind solche Angaben nur sehr schwach strukturiert, da hier keine Vorgaben hinsichtlich Inhalt und Format vorliegen. Diese vorhandenen Angaben können genutzt werden, indem die Suchmaschine sie als geographischen Bezug verwendet. Da dies heterogene Informationen ohne starke Struktur sind, sind sie häufig schwer zu identifizieren.

Google hat einen Ansatz entwickelt, dem dieses Konzept zugrunde liegt. Derzeit ist dieses Verfahren jedoch nicht öffentlich zugänglich.<sup>20</sup>

Stellvertretend für verschiedene Ortsangaben auf Web-Seiten werden hier zwei Beispiele für die am häufigsten vorkommende Ortsangabe, die Kontaktadresse, gegeben. Meist ist sie im Klartext eines HTML-Dokuments geschrieben, wie auch die Kontaktadresse auf der Web-Site des Fraunhofer Instituts für System und Softwaretechnik in Dortmund ([www.isst.fraunhofer.de](http://www.isst.fraunhofer.de)).

```
<p>Ansprechpartner für den Institutsteil Dortmund ist:</p>
<p>Dipl.-Volkswirt Carsten Lienemann<br>
  Emil-Figge-Str. 91<br>
  44227 Dortmund<br>
  E-Mail: <a href="mailto:info@do.isst.fhg.de">
           info@do.isst.fhg.de</a> <br>
<br>
```

Die in diesem Code implementierte Adresse des Ansprechpartners wird in einem Browser wie folgt dargestellt:

<sup>20</sup> Google (2002), <http://www.google.com/programming-contest/winner.html>.

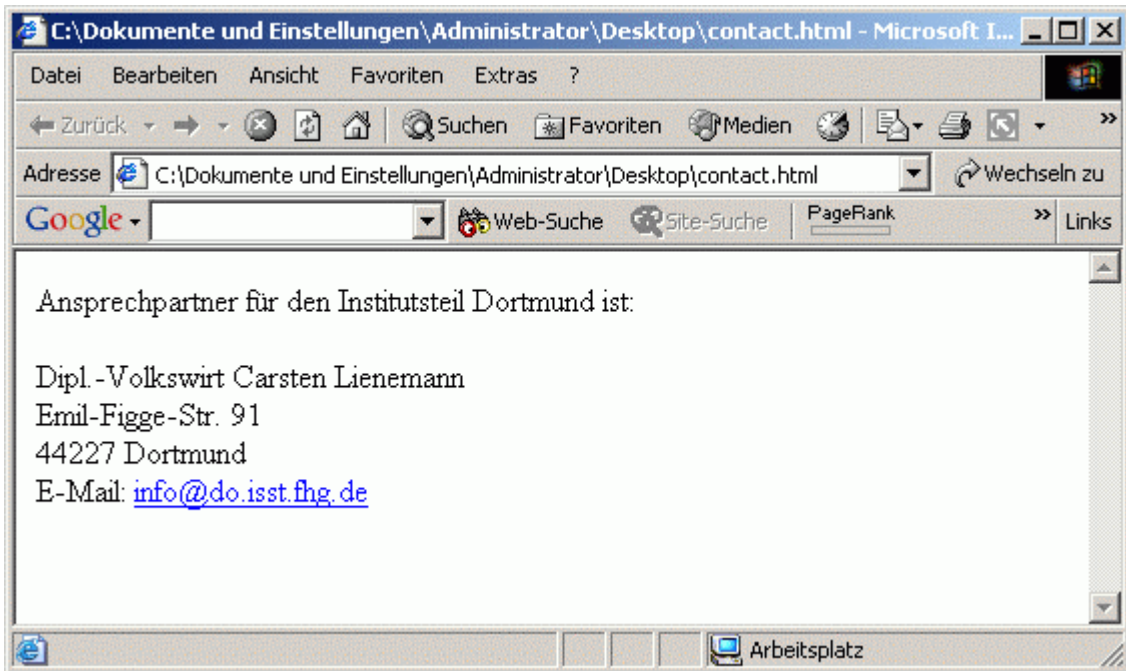


Abb. 4.3: Kontaktadresse in HTML

In dieser Form gespeicherte Adressen können nur aufgrund ihrer Struktur als solche erkannt werden. Wie im folgenden Beispiel zu erkennen ist, variiert diese Struktur in den verschiedenen Ländern sehr stark.

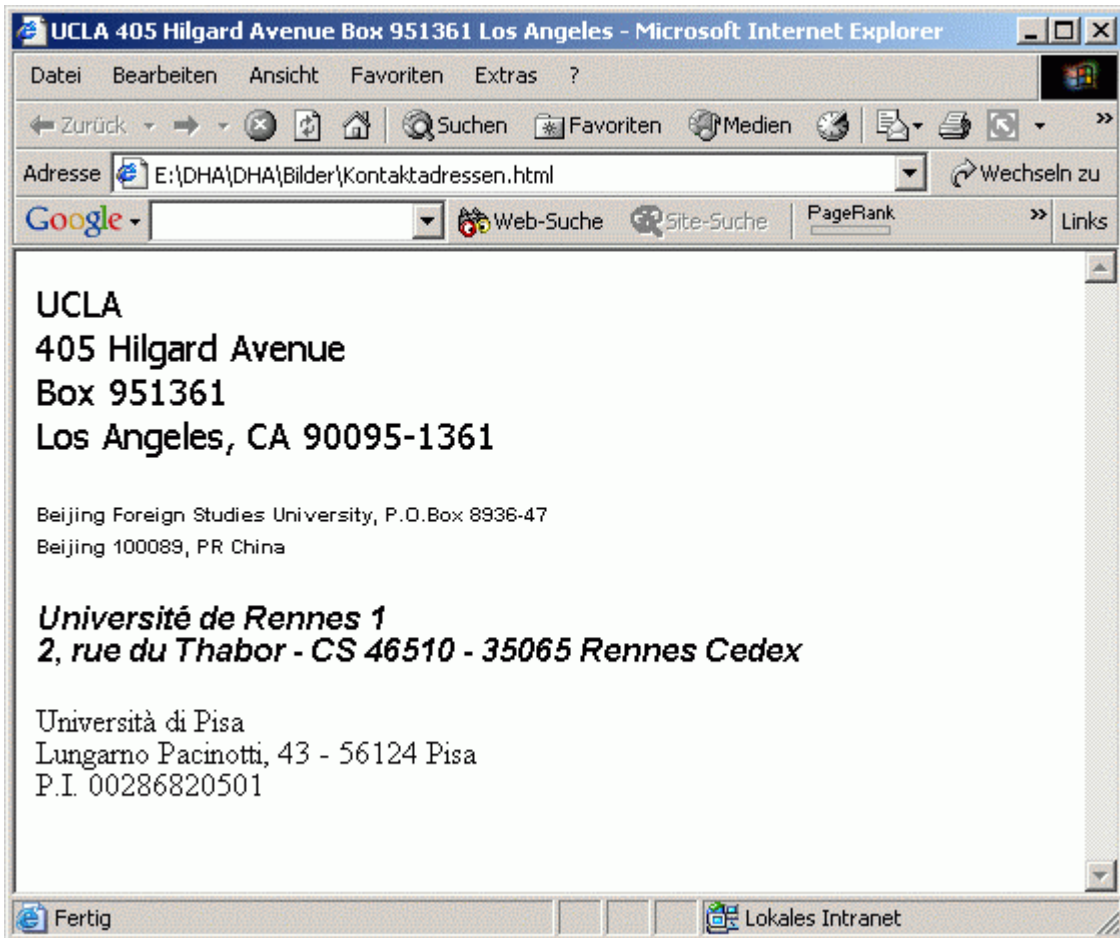


Abb. 4.4: Verschiedene Kontaktadressen

Obwohl alle in Abb. 4.4 dargestellten Adressen Kontaktadressen von Universitäten verschiedener Länder sind, sind sie sehr unterschiedlich bezüglich der Adressbestandteile und ihrer Reihenfolge. Daher ist es schwierig, die entsprechenden Textstellen, die diese Struktur aufweisen, zu erkennen und als geographischen Bezug zu speichern.

Adressangaben können auch in dem <address>-Tag, das in der HTML-Spezifikation 4.01 zur Verfügung gestellt wird, codiert werden. Hier wird die Adressangabe durch die Tags <address> und </address> umschlossen. Eine Adressangabe des Formats:

```

<ADDRESS>
<A href="../../../People/Raggett/">Dave Raggett</A>,
<A href="../../../People/Arnaud/">Arnaud Le Hors</A>,
contact persons for the <A href="Activity">W3C HTML Activity</A><BR>
$Date: 1999/12/24 23:07:14 $
</ADDRESS>21

```

wird von einem Browser wie folgt interpretiert:

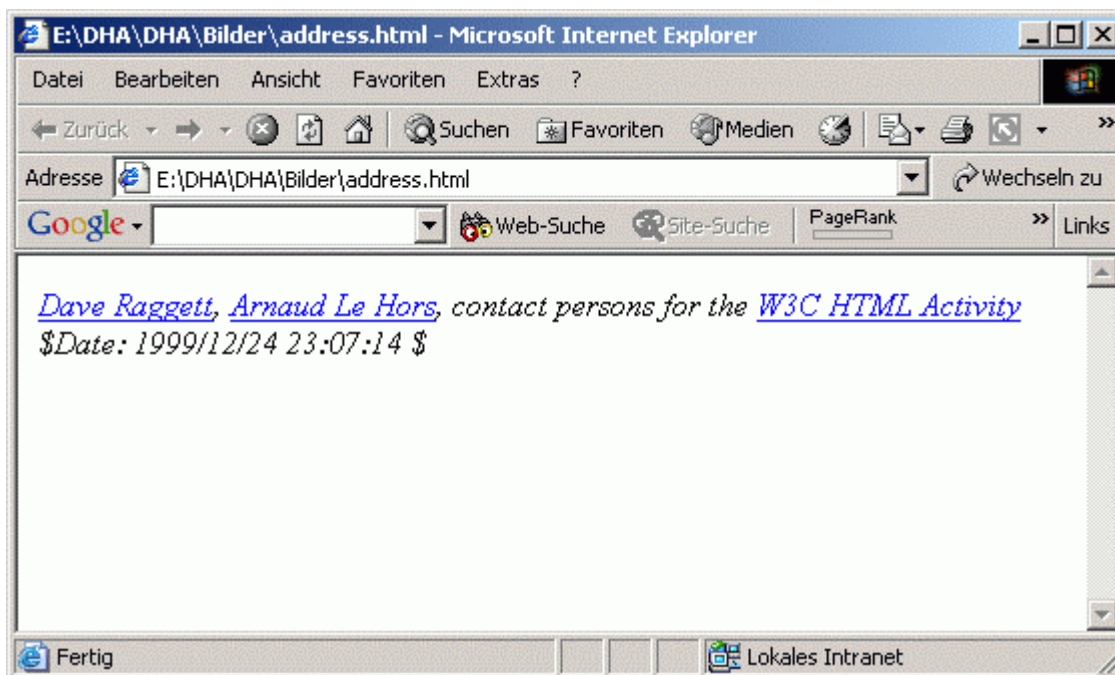


Abb. 4.5: Kontaktadresse innerhalb eines <address>-Tags

Die zwei Beispiele zeigen, dass neben den variierenden Formaten auch die Inhalte von Kontaktadressen unterschiedlich sind: sowohl die Straßenadresse als auch eine Emailadresse kann unter Kontaktangaben verstanden werden, aber nur die Straßenadresse weist auf einen geographischen Punkt dieser Erde hin. Den beiden Beispielen ist gemeinsam, dass die einzelnen Bestandteile wie Name, Straße u.a. nicht explizit als solche definiert wurden und somit nicht maschineninterpretierbar sind.

#### Web-Server-Standort

Außer Kontaktadressen und Anfahrtsbeschreibungen können auch andere Daten zur automatische Informationserhebung genutzt werden. Die Wahl des Server-Standortes kann Hinweise auf den geographischen Bezug der Web-Site enthalten. Eine Suchmaschine kann den Standort über das Länderkürzel oder über IP-Verfolgung identifizieren.

<sup>21</sup> HTML 4.01 Specification, Kapitel 7.5.6, <http://www.w3.org/TR/html401/struct/global.html#h-7.5.6>.

## Vergleich der Repräsentationskonzepte

Die dargestellten Möglichkeiten werden nun anhand der aufgestellten Anforderungen bewertet.

Verwendet ein *Registrierungsdienst den ISO-19115-Standard* als Grundlage für seine Eingabefelder, sind diese durch den Standard eindeutig definiert und bieten mächtige und leistungsstarke Konzepte. Fehlinterpretationen eines geographischen Bezugs kommen nicht vor. Jedoch sind solche Standards sehr umfangreich. Es bedarf einer langen Einarbeitungszeit in den ISO-Standard, um das Formular korrekt ausfüllen zu können, was nur professionelle Web-Site-Ersteller akzeptieren würden. Die Motivation des einfachen Autors wäre zu gering und damit scheitert dieses Konzept an mangelnder Beteiligung im Web.

Wird der geographische Bezug in eigens für diesen Zweck definierte *Metadaten* eingetragen, kann der Entwickler die Anzahl und Komplexität der anzugebenden Informationen festlegen und kommt so den Bedürfnissen des Web-Site-Autors entgegen, der ein einfaches und schnell durchschaubares Verfahren verlangt. Es muss jedoch dafür gesorgt werden, dass neben der Einfachheit die Eindeutigkeit der Informationen nicht verloren geht. Der Inhalt eines geographischen Bezugs muss genau festgelegt sein.

Die Verwendung von bereits *vorhandenen Geoinformationen in HTML-Dateien* hat den Vorteil, dass keine neuen Anforderungen auf die Web-Site-Autoren zukommen und bereits jetzt schon eine große Menge an Daten existieren. Probleme treten jedoch bei der Erkennung dieser schwach strukturierten Daten als geographischer Bezug auf: werden die Ortsangaben als Text in HTML geschrieben, können sie nicht immer eindeutig als Adressen erkannt werden. Außerdem ist die Vorstellung vom Inhalt einer Kontaktadressen sehr individuell.

Die Lokalisierung einer Web-Site über den *Standort des Web-Servers* hat den Vorteil, dass Informationen genutzt werden, die schon verfügbar sind, und somit bereits ein großer Datenbestand existiert. Jedoch stimmen die Rückschlüsse, die aus dem Standort des Servers gezogen werden, nicht unbedingt mit dem tatsächlichen geographischen Bezug des Inhalts überein. Server werden häufig an fremde Firmen zur Verwaltung gegeben. Viele internationale Firmen hosten ihre nationalen Web-Sites zentral auf einem Rechner. Statt des Landeskürzels geben dann Unterverzeichnisse im Pfad die Zugehörigkeit zu einem Land an. Oft ist es auch nicht möglich, aus der URL den Standort zu bestimmen, da anstelle der Länderkennung viele URLs Kürzel verwenden, die die Art der Institution beschreiben, wie ".com" für kommerzielle Organisationen oder ".gov" für Regierungsstellen. Beschränkt man sich auf die Analyse der IP-Adresse oder der Landesken-

nung, werden in den oben genannten Situationen keine oder falsche Bezüge festgestellt. Wie bei Kontaktadressen und Anfahrsbeschreibungen ist auch bei der Analyse des Standorts die Durchschaubarkeit des geographischen Bezugs nicht immer gegeben, da die Informationen nicht für einen geographischen Bezug erstellt wurden.

Des Weiteren können Web-Sites anhand von Landeskenntung oder IP-Adresse höchstens auf regionaler Ebene lokalisiert werden, für eine geographische Suche wird jedoch eine deutlich höhere Granularität erwartet. Daher ist diese Art der Lokalisierung zu ungenau.

### **4.3 Informationsakquisitionskomponente der Suchmaschine**

Der dritte Teil der Informationserhebung (s. Abb. 4.1) findet auf der Seite der Suchmaschine statt. Diese muss nun die im Web gespeicherten Informationen auslesen.

#### **Anforderungen an die Informationsakquisition der Suchmaschine**

Aus dem Bereich des Software Engineering werden folgende Anforderungen an die Informationsakquisition übertragen. Sie muss

- skalierbar sein. Die geographische Suchmaschine wird im Web eingesetzt, daher muss die Informationserhebung mit realistischem Aufwand für das ganze World Wide Web anwendbar sein.
- effizient sein, um in akzeptabler Geschwindigkeit Daten auswerten zu können. Eine annehmbare Performanz kann erreicht werden, wenn die Informationserhebung automatisch abläuft. Nur so können große Bestandteile des Webs analysiert werden.
- Daten korrekt analysieren. Die Fehlerquote bei der Analyse muss gering sein, um zuverlässige Ergebnisse der Suchmaschine sicherstellen zu können.
- flexibel sein - sowohl bezüglich der Geodatenformate, die es unterstützt als auch bezüglich teilweise fehlender Geodaten auf der Web-Site. Neue oder erweiterte Formate können entwickelt werden, um beispielsweise die Höhe als dritte Dimension oder komplexere Raumbezugsformen zu unterstützen.
- aktuell sein. Änderungen und Korrekturen von gespeicherten Geodaten müssen in der Datenbasis der Suchmaschine schnell durchführbar sein. Durch eine automatische Aktualisierung kann zusätzlich noch sichergestellt werden, dass dies in akzeptabler Zeit geschehen kann.

- wartbar sein. Wartung und Pflege darf auch bei einem großen Datenbestand nicht zu aufwändig werden, damit die Suchmaschine mit realistischem Aufwand für Datenbestände aus dem Web anwendbar ist.
- für unterschiedliche Netze und verschiedene Anwendungsfälle wiederverwendbar sein. Wenn die Daten getrennt von der Informationsakquisition der Suchmaschine gespeichert sind, können auch die Daten von anderen Diensten genutzt werden.

### Konzepte für die Informationsakquisition

Wie Informationen, die im Web verfügbar gemacht wurden, nun von der Suchmaschine extrahiert werden können, hängt von der Repräsentationsform im Web ab.

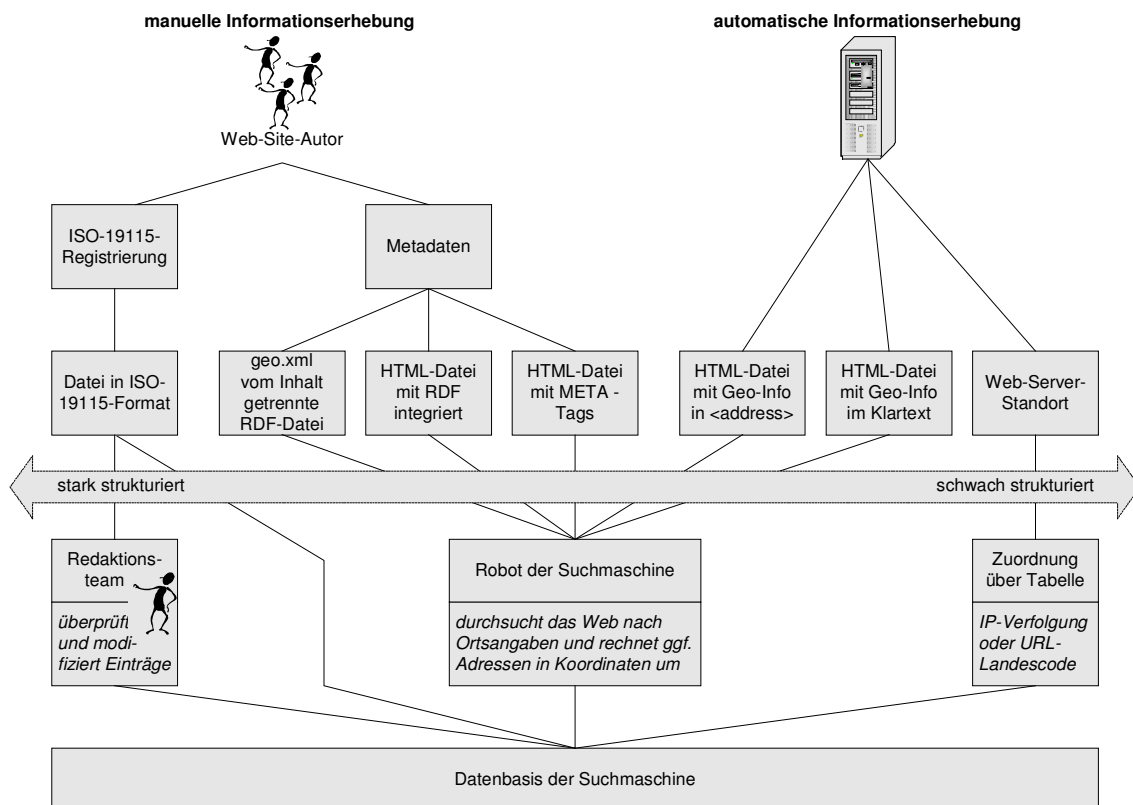


Abb. 4.6: Informationserhebung im Internet und -akquisition der Suchmaschine

#### Registrierungsdienst mit Redaktionsteam

Die Daten, die durch einen Registrierungsdienst erhoben wurden, können entweder direkt in die Datenbasis der Suchmaschine eingetragen werden oder vor der Veröffentlichung durch ein Redaktionsteam auf ihre Richtigkeit überprüft und gegebenenfalls modifiziert werden. Durch den Umfang und die Komplexität der ISO-19115 ist eine Kontrolle sinnvoll, um eine hohe Qualität der Daten zu garantieren.

Einige Suchmaschinen setzen ein Redaktionsteam zur Prüfung der Einträge ein. So wird bei der Suchmaschine [www.ms-guide.de](http://www.ms-guide.de) der Firma [unARTIG.de](http://unARTIG.de) jeder Eintrag redaktionell überprüft, bevor er in die Suchmaschine eingepflegt wird.

#### *Robot einer Suchmaschine*

Wird ein Verfahren gewählt, in dem die geographischen Informationen explizit durch ein vorgeschriebenes Format auf der Homepage oder implizit über schwach strukturierte Informationen hinterlegt werden, wird ein Robot einer Suchmaschine benötigt, der das gesamte Web kontinuierlich durchsucht. Seine Aufgabe ist es, die gesuchten Stellen zu finden und mit dem Verweis auf die Web-Site in der Suchmaschine zu speichern. Ändert sich der geographische Bezug einer Web-Site bzw. Web-Seite, so werden die Daten innerhalb der Homepage geändert. Wird die Web-Site von dem Robot erneut gescannt, kann auch der Eintrag in der Suchmaschine aktualisiert werden.

Da Daten vorliegen, die sich in Format und Inhalt sehr unterscheiden, besteht eine weitere Anforderung darin, die Daten trotz verschiedener Struktur zu erkennen. Eine Ortsangabe in Adressform muss über einen Geo-Coder oder einen Gazetteer in Koordinaten umgerechnet werden. Um eine möglichst einheitliche Umrechnung in Koordinaten zu erhalten, wird empfohlen, einen Referenz-Geo-Coder festzulegen, in dem beispielsweise das Stadtgebiet Münster eindeutig einem Polygon zugewiesen ist.

#### *Zuordnung des Web-Server-Standorts*

Die Zuordnung des Inhaltes einer Web-Site kann beispielsweise über den physischen Ort des Web-Servers geschehen. Die IP-Adresse und das Landeskürzel des Domain-Namens weisen auf mögliche Korrelationen hin.

Um zu einer IP-Adresse den zugehörigen Domain-Namen zu ermitteln, kann der DNS-Lookup-Dienst genutzt werden. Über die Landeskennung des Domain-Namens kann die Suchmaschine bereits erste Informationen über den Standort des Web-Servers erhalten. Ein WWW-Whois-Dienst liefert Informationen über den Domain-Registranten.<sup>22</sup> Abb. 4.7 zeigt Teile des Eintrags der Domain [uni-muenster.de](http://uni-muenster.de). Die vollständige Ausgabe des Dienstes ist im Anhang zu finden.

---

<sup>22</sup> Vgl. Röscheisen, E. (2002), S. 90 ff.

```

uni-muenster.de
Request: uni-muenster.de

domain:      uni-muenster.de
descr:      Westfaelische Wilhelms-Universitaet
descr:      Zentrum fuer Informationsverarbeitung
descr:      Roentgenstrasse 9-13
descr:      48149 Muenster
descr:      Germany
....

```

Abb. 4.7: Informationen eines WWW-Whois-Dienstes<sup>23</sup>

In diesem Fall lässt sich der physische Standort des Servers eindeutig bestimmen und durch die Ortsangabe Münster wurde auch ein korrekter geographischer Bezug des Inhalts der Web-Site gefunden. Dies trifft aber nur dann zu, wenn die geographische Referenz mit der Ortsangabe des Domain-Registranten übereinstimmt. Ansätze, das Domain Name System (DNS) durch Ortsinformationen zu erweitern, haben sich bisher nicht durchgesetzt. In dem "DNS LOC Resource Record" sollen geographische Informationen über den Server gespeichert werden.<sup>24</sup>

```

OrgName:      Westfaelische Wilhelms - Universitaet
OrgID:        WWU-1

NetRange:     128.176.0.0 - 128.176.255.255

...

TechHandle:   GR26-ARIN
TechName:     Richter, Georg
TechPhone:    +49 251 83 31611
TechEmail:    network@uni-muenster.de

```

Abb. 4.8: Informationen eines IP-Whois-Dienstes<sup>25</sup>

Ist der IP-Nummer kein Domain-Name zugeordnet, kann auch mit der IP-Adresse eine Lookup-Anfrage an einen IP-Whois-Dienst gestellt werden. So erhält die geographische Suchmaschine Informationen über den Registranten des entsprechenden Teilnetz-Adressblocks, aus denen wie aus der Adresse des Domain-Registranten Rückschlüsse

<sup>23</sup> Diese Informationen sind von dem Dienst ALLWHOIS entnommen, <http://www.allwhois.com>.

<sup>24</sup> Vgl. Davis, C. (2001).

<sup>25</sup> Diese Informationen sind von dem Dienst ip-telligence entnommen, <http://ailab.de>.

auf den geographischen Bezug der Web-Site gezogen werden können. In dem oben genannten Beispiel (s. Abb. 4.8) kann die Telefonnummer den Hinweis auf Münster geben.<sup>26</sup>

Alternativ zu Whois-Diensten kann die geographische Lokalisierung der Web-Site auch anhand einer Localizer-Datenbank, der NetGeo-Datenbank oder des Traceroutes geschehen. Die Localizer-Datenbank hat Ortsangaben von ca. 75% aller IP-Adressen im deutschsprachigen Raum verzeichnet. Mittels Traceroute wird der letzte Router vor Erreichen des gewünschten Servers gesucht. Da die meisten Provider ihren Routern ortsbezogene Namen geben, können diese als Hinweis für den geographischen Standort des Servers verwendet werden. Bei der Suche nach der IP des Servers der Universität Münster wird jedoch die Stadt Bielefeld angegeben.<sup>27</sup>

	C-API
City	Bielefeld
Province	Nordrhein-Westfalen
Country	Germany
(Backbone-) Provider	Deutsches Forschungsnetz
Picture	geological position
Elapsed time for	
	C-API
loading the database	5.10ms
the lookup	0.31ms
SHM Support	enabled
SHM size (in kbyte)	2366.84
entries in DB	75083
strings in DB	1173

Abb. 4.9: Informationen der Localizer-Datenbank<sup>28</sup>

Die geographische Internet-Datenbank NetGeo der Cooperative Association for Internet Data Analysis (CAIDA) analysiert die Einträge einer Whois-Datenbank, um den Ortsbezug einer Web-Site festzustellen. Durch ein Analysetool ist CAIDA in der Lage, basierend auf der NetGeo-Datenbank ein Traceroute graphisch anzuzeigen. Wie die Whois-Dienste verwendet die NetGeo Datenbank die Adresse des Registranten, um einen Ortsbezug für die Web-Site festzulegen.<sup>29</sup>

<sup>26</sup> Vgl. Röscheisen, E. (2002), S. 90 ff.

<sup>27</sup> Vgl. Röscheisen, E. (2002), S. 90 ff.

<sup>28</sup> Diese Information stammt aus der Localizer-Datenbank, <http://jan.kneschke.de/projects/localizer>.

<sup>29</sup> Vgl. Röscheisen, E. (2002), S. 90 ff.

Diese Methoden werden bereits erfolgreich angewendet, um Internet-Surfer zu lokalisieren. Mit dieser Information kann beispielsweise dynamisch lokale Werbung auf die Internetseiten geschaltet werden.

Das Verfahren, über die Landeskenntung den geographischen Bezug einer Web-Site festzustellen, wird schon in verschiedenen Suchmaschinen angewendet. Als Beispiel sei hier die Suchmaschine HOTBOT von Lycos genannt. In der erweiterten Suchfunktionalität der Suchmaschine kann die Region bzw. die Domain einer Web-Site als zusätzliches Suchkriterium eingesetzt werden.<sup>30</sup>

### **Vergleich der Informationsakquisitionskonzepte**

Wenn der *Registrierungsdienst mit Redaktionsteam* zentral gehalten und jeder Eintrag durch ein Redaktionsteam geprüft wird, ist dieses Verfahren nur in beschränktem Maße skalierbar und verliert bei wachsender Datenmenge zu schnell an Performanz und Effizienz. Auf das Web ist diese Methode daher nicht anwendbar. Vorteilhaft ist, dass Daten durch ihre fest vorgegebene Struktur von der Suchmaschine fehlerfrei analysierbar sind. Neue Datenformate müssen auf der Ebene des Registrierungsdienstes eingepflegt werden und sind für die Informationsakquisition der Suchmaschine unproblematisch. Bei Änderungen des geographischen Bezugs müssen die Daten vom Web-Site-Autor aktualisiert und neu in der Suchmaschine abgespeichert werden. Dieses Konzept ist gut wartbar und auch für andere Anwendungsbereiche wiederverwendbar. Nachteilig jedoch wirkt sich aus, dass die Daten selbst nur in der Suchmaschine vorhanden sind und somit nicht von anderen Diensten genutzt werden können.

Der Einsatz eines *Robot* hat sich bei der Informationsakquisition für Suchmaschinen in der Praxis bewährt. Jedoch können nur qualitativ hochwertige, d.h. korrekte Daten, die eine ausreichende geographische Genauigkeit beinhalten, extrahiert werden. Diese müssen in einer maschinenlesbaren, idealer Weise in einer maschineninterpretierbaren Form vorhanden sein. Unstrukturierte bzw. schwach strukturierte Daten sind nur schwer zu erkennen und richtig zu deuten. Somit besteht hier die Gefahr, durch eine zu hohe Fehlerquote die Suchergebnisse zu verfälschen. Bei Änderung des geographischen Bezugs muss der Web-Site-Autor die Daten auf seiner Web-Site korrigieren, damit diese bei einem erneuten Durchlauf der Suchmaschine aktualisiert werden können.

---

<sup>30</sup> Weitere Informationen über die Suchfunktionalitäten von Lycos Hotbot sind unter <http://hotbot.lycos.com/?query=&cobrand=&matchmode=all&datedelta=0&language=any&recordcount=10&descriptiontype=2&modsign1=MC&dateoption=within&placeselection=georegion&act.super.x=82&act.super.y=14> zu finden.

Dieses Verfahren eignet sich für den Einsatz im Web, falls die Daten in einem interpretierbaren Format vorliegen. Da die Informationen auf der Web-Site vorhanden sind, können sie von anderen Diensten oder Suchmaschinen ebenfalls verwendet werden.

Sowohl über Länderkürzel als auch über die IP-Adresse lässt sich der *Standort des Web-Servers* einem geographischen Bezug zuordnen. Diese Zuordnung ist für das gesamte Web effizient realisierbar. Die Länderkürzel sind eindeutig und durch eine Suchmaschine korrekt analysierbar. Viele IP-Adressen können auch zuverlässig interpretiert werden. Änderungen der IP-Adresse und des Länderkürzels werden durch die kontinuierlichen Suchdurchläufe der Suchmaschine regelmäßig überprüft. Für eine geographische Suche kann dieses Konzept jedoch nicht verwendet werden, da sowohl das Länderkürzel als auch die IP-Adresse auf höherer Granularitätsstufe keine eindeutige Struktur mehr aufweisen und somit nicht zuverlässig gedeutet werden können. Auch ist die Lokalisierung der Web-Server sehr unsicher und der hergeleitete Ortsbezug entspricht oft nicht dem geographischen Bezug des Web-Site-Inhalts. Der Web-Server-Standort eines Touristeninformationszentrums stimmt beispielsweise selten mit dem geographischen Standort der dort beschriebenen Sehenswürdigkeiten überein. Gleiches gilt auch für die implizit implementierten Kontaktadressen oder Anfahrtsbeschreibungen.

#### **4.4 Informationsbasis der Suchmaschine**

Die während der Informationserhebung erhaltenen Daten müssen in der Informationsbasis einer Suchmaschine strukturiert gespeichert werden, damit Abfragen auf dem Datenbestand durchgeführt und die Ergebnisse visualisiert werden können.

Um Geodaten in der Informationsbasis effizient verarbeiten zu können, muss diese den spezifischen Eigenschaften räumlicher Daten angepasst sein. Allgemeine Anforderungen an die Datenspeicherung von Suchmaschinen werden hier nicht näher untersucht.

##### **Eigenschaften räumlicher Daten**

Räumliche und nicht-räumliche Daten unterscheiden sich wesentlich: nicht-räumliche Daten bestehen tendenziell aus einfachen und strukturierten Datensätzen von geringer Größe. Die Beziehungen zwischen den Daten sind eher gering und die Transaktionen kurz. Räumliche Daten hingegen sind häufig mehrdimensional und enthalten sowohl Informationen über das Objekt als auch über seine Position im Raum. Einige grundlegende Eigenschaften werden im Folgenden aufgelistet:

- Räumliche Datenobjekte haben oft eine sehr komplexe Struktur - sie können beispielsweise nur aus einem Punkt bestehen, aber auch aus vielen tausend Polygonen, die willkürlich im Raum verteilt sind.
- Räumliche Datensätze sind häufig sehr groß. Geographische Karten benötigen typischerweise mehrere Gigabyte Speicher.
- Für räumliche Daten gibt es keine Standardalgebra. Auch wenn keine festgelegten Basisoperatoren für räumliche Daten existieren, muss die Datenbasis alle für den Anwendungsfall benötigten Funktionen unterstützen.
- Viele räumliche Operatoren sind nicht abgeschlossen. Die Rückgabe einer Schnittmenge von zwei Polygonen kann sowohl eine Menge an Punkten, verschiedene Kanten oder disjunkte Polygone sein. Die Datenbasis muss solche Operatoren ausführen und das Ergebnis speichern können.
- Räumliche Operatoren sind in der Regel aufwändiger und damit teurer als relationale Operatoren. Trotzdem darf die Performanz und die Effizienz nicht zu stark sinken.<sup>31</sup>

Im Gegensatz zur herkömmlichen Datenbasen muss eine räumliche Datenbasis zwei wesentliche Anforderungen erfüllen: sie muss sowohl räumliche Datentypen und als auch räumliche Operationen unterstützen.

#### **4.5 Abfragekomponente der Suchmaschine**

Neben der Datenbasis muss auch die Abfragekomponente einer geographischen Suchmaschine räumliche Daten unterstützen und sowohl raumbezogene als auch die nicht-raumbezogene Abfragen verarbeiten. Im Folgenden werden die Eigenschaften einer raumbezogenen Suchanfrage, räumlicher Datentypen und räumlicher Operationen untersucht. Anschließend werden Anforderungen definiert und verschiedene Konzepte vorgeschlagen, die in der Diskussion gegenübergestellt und anhand der Anforderungen bewertet werden. In den Abfragen selektiert der Anwender häufig nicht nur Daten, sondern legt auch die graphische Ausgestaltung des Ergebnisses fest. Auf die Darstellungsmöglichkeiten wird im folgenden Kapitel eingegangen.<sup>32</sup>

---

<sup>31</sup> Vgl. Günther, O. (1998), S. 50 ff.

<sup>32</sup> Vgl. Bill, R. (1996), S. 348.

## Eigenschaften raumbezogener Abfragen, Datentypen und Operationen

Beim Abrufen von Geoinformationen aus der Informationsbasis entstehen durch den Raumbezug typische Abfragesituationen (s. Abb. 4.10), von denen einige beispielhaft aufgeführt werden:

- "Die Suche eines Punktes über seine Koordinaten, wie sie etwa beim interaktiven Identifizieren über das Fadenkreuz am Bildschirm notwendig wird; meist ist nur eine genäherte Position bekannt.
- Das interaktive Identifizieren eines Objektes am Bildschirm, das auf die Identifikation entsprechender Objektpunkte zurückgeführt werden kann.

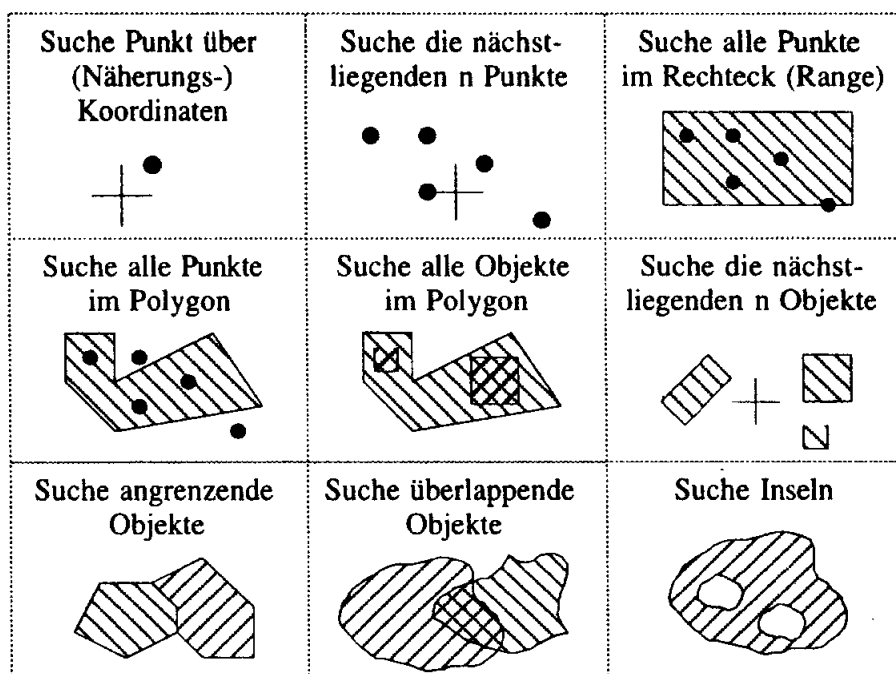


Abb. 4.10: Raumbezogene Suchanfragen<sup>33</sup>

- Die Suche aller Punkte oder Objekte in einem Rechteck [...]; dies entspricht einer Ausschnittsbildung, einem Zoom oder einer Bildschirmauffrischung. Auch Objektteile, die in das Rechteck hereinragen, sind zu ermitteln und gegebenenfalls abzuschneiden.
- Die Suche aller Punkte oder Objekte in einem allgemeinen Polygon; sie lässt sich zunächst näherungsweise auf eine Suche im umschreibenden Rechteck reduzieren

<sup>33</sup> Bartelme (2000), S. 250.

[...], wonach in einem zweiten Schritt die Aussonderung jener Punkte bzw. Objekte vorzunehmen ist, die zwar im Rechteck, nicht aber im Polygon liegen [...].

- Die Suche der nächstliegenden n Punkte oder Objekte, sortiert gemäß ihrer Entfernung vom vorgegebenen Zentrum (Dies ist grundsätzlich nicht durch die Vorgabe eines Fensters lösbar, wenn man nichts über die Verteilung der Daten weiß.)
- Die Suche nach benachbarten Objekten zu einem vorgegebenen Objekt oder nach Objekten, die dieses schneiden, enthalten oder von diesem gedeckt werden bzw. eine Aussparung (Insel) zu diesem Objekt bilden."<sup>34</sup>

Um diese raumbezogenen Abfragen durchführen zu können, werden räumliche Datentypen benötigt. Einem räumlichen Objekt ist genau eine eindeutige Stelle in einem d-dimensionalen euklidischen Raum oder einem Unterraum von diesem zuzuordnen. Ein räumlicher Datentyp muss diese Eigenschaft geeignet abbilden können. Beispiele für in der Suchmaschine verwendete Datentypen sind Punkte und Linien, die ein eindimensionales Polyhedron kennzeichnen, Polygone oder Regionen, die ein zweidimensionales Polyhedron bezeichnen oder seltener auch Würfel, die ein dreidimensionales Polyhedron kennzeichnen.<sup>35</sup>

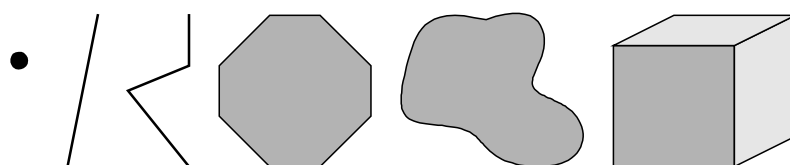


Abb. 4.11: Räumliche Datentypen

Um die Geodaten zu verarbeiten und Analysen auf ihnen durchzuführen, werden räumliche Operatoren benötigt, die auf räumlichen Datentypen arbeiten. Einige typische sind in den folgenden Abbildungen zusammengestellt.

Um Nachbarschaftsbeziehungen zweier Objekte zu untersuchen, werden Operatoren zur Berechnung von Abstand und Länge einer Strecke benötigt. Für die Suche eines Punktes über seine Näherungskordinaten wird der Abstandsoperator verwendet. Um den optimalen Treffpunkt zweier Reisenden zu berechnen, muss der Schnittpunkt ihrer Wege berechnet werden.

<sup>34</sup> Bartelme, N. (2000), S. 250, 251.

<sup>35</sup> Vgl. Günther, O. (1998), S. 51-53.

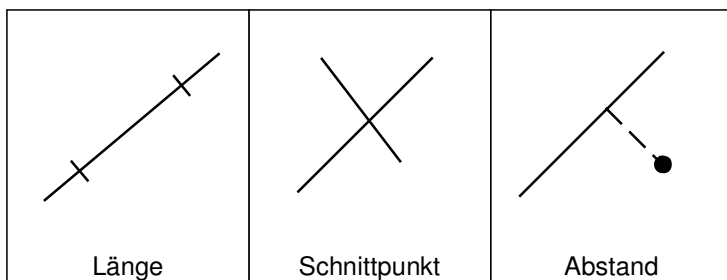


Abb. 4.12: Räumliche Operatoren I<sup>36</sup>

Um Suchergebnisse nach geographischen Kriterien zu sortieren, kann die Größe des geographischen Bezugs herangezogen werden, die über die Flächenberechnung bestimmt wird. Zur Suche aller Punkte oder Objekte in einer Fläche wird der Inklusionsoperator benötigt.

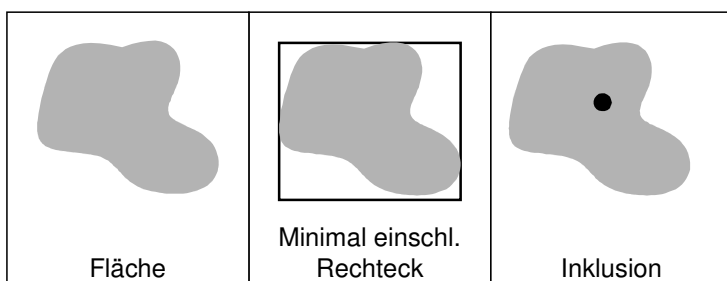


Abb. 4.13: Räumliche Operatoren II<sup>37</sup>

Überlappende Objekte oder Inseln können mit räumlichen Operatoren wie Vereinigung, Durchschnitt und Differenz festgestellt werden.

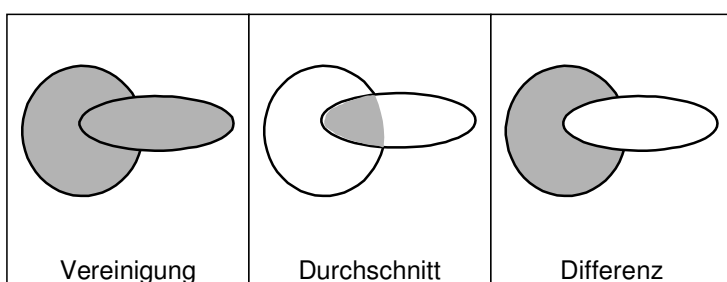


Abb. 4.14: Räumliche Operatoren III<sup>38</sup>

Diese vom Nutzer gestellten Anfragen muss die Abfragekomponente der Suchmaschine in eine formale Sprache übersetzen. Durch den Raumbezug sind folgende Eigenschaften für eine raumbezogene Abfragesprache unerlässlich:

<sup>36</sup> Vgl. Bill, R. (1996), S. 346.

<sup>37</sup> Vgl. Bill, R. (1996), S. 346.

<sup>38</sup> Vgl. Bill, R. (1996), S. 346.

- Die Möglichkeit, in der Abfragesprache geometrische, topologische und attributive Einschränkungen vornehmen zu können.
- Die Möglichkeit, in der Abfragesprache die graphische Ausgestaltung und Darstellungsform des Ergebnisses zu bestimmen.<sup>39</sup>

### **Konzepte für raumbezogene Abfragesprachen**

Es werden drei Abfragemethoden unterschieden: Dialogsysteme, natürliche und formale Abfragesprachen. Formale Abfragesprachen können weiter unterteilt werden in eigenständig entwickelte Sprachen und bezüglich des Raumbezugs erweiterte Sprachen.

#### *Dialogsysteme*

Dialogsysteme erfragen jeden einzelnen Schritt vom Benutzer. Muss eine Entscheidung über die weitere Bearbeitungsfolge getroffen werden, zeigt das System dem Anwender sämtliche Alternativen an, aus denen dieser dann auswählen kann. Viele Dialogsysteme sind in der Lage, nicht nur abhängig von der Benutzereingabe, sondern auch vom Inhalt der Datenbasis ihre Bearbeitungsfolge zu modifizieren.<sup>40</sup>

#### *Natürliche Abfragesprachen*

Über natürliche Abfragesprachen wird die Kommunikation zwischen Mensch und Maschine in einer natürlichen, weitgehend formatfreien Sprache realisiert. Der Mensch drückt seine Wünsche in individuellen Sätzen aus. Diese gilt es durch eine Abfragesprache mit einem umfangreichen Wortschatz maschinenlesbar und -interpretierbar zu machen. Jedoch führt die Formatfreiheit dieser natürlichen Sprache dazu, dass besonders die komplexen raumbezogenen Abfragen schwer auszuführen sind: komplizierte Kombinationen von Selektionskriterien und Auswertungen bedürfen einer exakten Formulierung, um logische Fehler zu vermeiden.<sup>41</sup>

#### *Formale Abfragesprachen*

Im GIS-Bereich werden derzeit vor allem formale Abfragesprachen verwendet. Solche Sprachen basieren auf einer englisch-ähnlichen Schlüsselwortsyntax verknüpft mit einer

---

<sup>39</sup> Vgl. Egenhofer, M. J. (1990).

<sup>40</sup> Vgl. Bill, R. (1996), S. 347.

Vgl. Findeisen, D. (1990), S. 107-109.

<sup>41</sup> Vgl. Bill, R. (1996), S. 347.

Vgl. Findeisen, D. (1990), S. 115-116.

mathematischen Symbolik. Durch ihre eindeutige und klare Definition ist dies eine schnelle und flexible Methode, um Daten abzufragen. Da solche Sprachen häufig aus gekürzten Schlüsselwörtern und vielen arithmetischen Operatoren bestehen, sind sie für den Laien relativ schwer zu erlernen und zu handhaben. Durch zahlreiche Kombinationsmöglichkeiten von Selektionskriterien und Operationen erreichen formale Abfragesprachen eine sehr hohe Leistungs- und Einsatzbreite.<sup>42</sup>

Vielen herkömmlichen formalen Abfragesprachen fehlten Selektionskriterien und Sprachelemente, um topologische Beziehungen zu identifizieren und abzufragen. Daher entstanden verschiedene raumbezogene Abfragesprachen entweder als neue, eigenständige Sprachen oder als Sprachen, deren Entwicklung sich an existierenden, erfolgreichen Sprachen orientiert.

#### *Erweiterung existierender Abfragesprachen*

Viele raumbezogene Abfragesprachen wurden durch Erweiterung von existierenden, meist weit verbreiteten und akzeptierten Sprachen entwickelt. Die bekannteste Abfragesprache SQL wurde wohl am häufigsten für Weiterentwicklungen herangezogen. Zentrale Erweiterungen sind die Einbettung räumlicher Datentypen, räumlicher Operatoren und graphischer Ausgabemöglichkeiten bzw. die flexible Unterstützung eigener Datentypen und darauf anzuwendende Operatoren.

#### *Entwurf einer eigenständigen, neuen Sprache*

Eigenständige Sprachen werden meistens in Kombination mit Produkten entwickelt, es gibt jedoch auch einige allgemeine, produktneutrale Entwürfe wie MAP-Query von A.U. Frank, LOBSTER von M. J. Egenhofer und GEOQL von M. S. Bundock. Wichtige Anforderungen an raumbezogene Abfragesprachen sind die Integration der graphischen Ausgabe. Bei der Abfrage wird auch die Präsentation der Ergebnisse festgelegt, beispielsweise durch Operatoren wie Überlagerung, Hervorheben, Erneuern.<sup>43</sup>

### **Vergleich der Konzepte für raumbezogene Abfragesprachen**

Um *Dialogsysteme* zu benutzen sind weder Detailkenntnisse vom Abfrageprogramm noch von dem Dateninhalt erforderlich. Dadurch ist dieses Verfahren für den Anwender sehr einfach zu erlernen und leicht zu handhaben. Allerdings ist die Effizienz und Ge-

---

<sup>42</sup> Vgl. Bill, R. (1996), S. 347.

Vgl. Findeisen, D. (1990), S. 109-115.

<sup>43</sup> Bill, R. (1996), S. 345.

schwindigkeit von Dialogsystemen sehr gering, weil sich zu den Wartezeiten, die aus den Datenauswertungen resultieren, die Bearbeitungszeiten des Benutzers addieren. Da Dialogsysteme stark abhängig von dem zugrundeliegenden Datenmodell sind, sind die Abfragemöglichkeiten begrenzt und sehr starr. Komplexe Abfragen, wie sie für Geodaten erforderlich sind, werden daher eher schlecht unterstützt.

*Natürliche Sprachen* unterstützen den Menschen in seiner individuellen Ausdrucksweise. Dadurch jedoch verliert die Sprache die Möglichkeit, komplexe Anfragen exakt auszudrücken. Da Abfragen auf Geoinformationen häufig sehr kompliziert sind, ist diese Art der Abfrage für eine Geodatenbasis nicht besonders geeignet. Überdies ist nach wie vor die Implementierung und Realisierung einer natürlichen Sprache mit einem hohen Aufwand verbunden.

Die *formale Abfragesprache* ist eine schnelle und flexible Methode zur Formulierung von Datenauswertungen. Auch sehr komplizierte geographische Abfragen können mit formalen Abfragesprachen exakt ausgedrückt werden. Jedoch muss die sprachspezifische Syntax erlernt werden und kann daher für den Laien mit Problemen verbunden sein. Wegen ihrer verglichen mit den anderen Sprachkonzepten sehr hohen Leistungsfähigkeit wird sie häufig im GIS Bereich eingesetzt.

Die Erweiterung einer *existierenden Abfragesprache* um räumliche Datentypen hat den Vorteil, dass eine Sprache erweitert werden kann, bei der es sich schon um einen weit verbreiteten Standard handelt. So wurde häufig SQL um raumbezogene Abfragemöglichkeiten erweitert. Nachteilig wirkt sich aus, dass sich die erweiterte Sprache an die Gesamtstruktur der Originalsprache anlehnen muss und deshalb teilweise Kompromisse hinsichtlich des Raumbezugs eingegangen werden müssen.

Der *Entwurf einer eigenständigen Sprache* kann sich als ein Neuentwurf wirklich und ausschließlich auf die geeignete Modellierung und Abfrage raumbezogener Daten unabhängig von anderen Anwendung konzentrieren. Da eine solche Sprache an keinen Standard angelehnt ist, hat sie mit erheblichen Akzeptanzschwierigkeiten zu kämpfen.

#### **4.6 Präsentationskomponente der Suchmaschine**

Die Präsentationskomponente der Suchmaschine muss die Ergebnisse der Suchanfrage übersichtlich und für den Benutzer verständlich und lesbar darstellen. In herkömmlichen Suchmaschinen werden die Treffer textuell nach verschiedenen Kriterien sortiert aufgelistet. Eine geographische Suchmaschine unterstützt neben der textuellen Präsentation der Ergebnisse auch eine graphische Darstellung. Zur Präsentation geographischer Informationen wird die Visualisierung genutzt mit dem Ziel, die Analyse, das Verständnis

und die Kommunikation von Modellen, Konzepten und Daten in der Wissenschaft zu erleichtern.<sup>44</sup> Im Folgenden wird untersucht, welche Anforderungen an eine übersichtliche Präsentation gestellt werden.

### **Anforderungen an die Präsentation von Geodaten**

Die Präsentation von Geodaten bildet die Eigenschaften der Datenmenge auf geographische Elemente und deren Attribute ab. Die Qualität der Visualisierung wird anhand verschiedener Kriterien wie Expressivität, Effektivität und Angemessenheit beurteilt.

- Ziel einer graphischen Darstellung von Geodaten besteht darin, die Datenmenge möglichst unverfälscht wiederzugeben. Die Wahl einer Visualisierungstechnik beeinflusst die Ausdrucksfähigkeit einer Darstellung, die sogenannte Expressivität, elementar. Die Resultate einer geographischen Suche lassen sich am besten durch eine kartenverwandte Darstellung abbilden.<sup>45</sup>
- Häufig gibt es mehrere Darstellungstechniken, die sich zur Präsentation von Geodaten eignen. Um eine möglichst hohe Effektivität einer Darstellung zu erreichen, müssen die Fähigkeiten des Betrachters, die charakteristischen Eigenschaften des Ausgabegerätes sowie die Zielsetzung und der Anwendungskontext optimal ausgenutzt werden. Die Effektivität hängt also anders als die Expressivität nicht allein von den Daten ab. Besonders bei der Präsentation auf mobilen Endgeräten wird die Darstellungsart z.B. durch Bildschirmgröße, Farbfähigkeit und Datenübertragungsrate stark beeinflusst.<sup>46</sup>
- Da der Nutzen der Visualisierung auch mit Aufwand verbunden ist, muss die Darstellungsart neben der Expressivität und der Effektivität auch die Angemessenheit bezüglich Aufwand und Kosten berücksichtigen. Hier wird im Gegensatz zu den beiden ersten Anforderungen nicht die Visualisierung selbst sondern die Durchführung des Visualisierungsprozesses betrachtet.<sup>47</sup>

### **Konzepte für die Präsentation von Geodaten**

Für räumliche Daten existiert eine Vielzahl verschiedener Darstellungsmöglichkeiten. Hier werden nur diejenigen besprochen, die für die Darstellung von Suchergebnissen in Frage kommen.

---

<sup>44</sup> Vgl. Schumann, H.; Müller, W. (2000), S. 5.

<sup>45</sup> Vgl. Schumann, H.; Müller, W. (2000), S. 9-10.

<sup>46</sup> Vgl. Schumann, H.; Müller, W. (2000), S. 11-12.

<sup>47</sup> Vgl. Schumann, H.; Müller, W. (2000), S. 12-13.

### *Interaktive Graphik*

Die gängigste Darstellungsform von Geodaten ist die interaktive Graphik am Bildschirm in Form von kartenverwandten Darstellungen, z.B. thematische, interaktive Karten. An diese Schnittstelle werden besonders hohe Erwartungen hinsichtlich Leistungsfähigkeit, Geschwindigkeit, Benutzerfreundlichkeit und Flexibilität gestellt. Lange und stark streuende Wartezeiten führen zu Akzeptanzproblemen.

Neben den Suchanfragen können an eine interaktive Graphik auch Anfragen bezüglich der Präsentation gestellt werden. Wird eine Suchanfrage an die Suchmaschine gerichtet, ist der Anfrageraum die Informationsbasis. Bei Anfragen dagegen, die die Darstellungsform wie z.B. die Auflösung betreffen, ist der Anfrageraum die Graphik. Nicht alle gefundenen Treffer müssen bei der Präsentation auf dem Bildschirm angezeigt werden. Wird eine Anfrage für einen größeren als am Bildschirm angezeigten Kartenausschnitt gestellt, so werden auf der Graphik nur die Treffer innerhalb des Kartenausschnitts dargestellt.

### *Passive Graphik*

Statt der interaktiven kann auch eine passive Graphik als Darstellungsform gewählt werden. Wie oben wird wohl auch die passive Graphik am häufigsten durch eine kartenverwandte Darstellung realisiert, aber auch Bilder, Photos und Videos sind mögliche Darstellungsarten.

Nichtgraphische Darstellungsformen wie Auflistungen, Tabellen und Berichte werden in nicht-geographischen Suchmaschinen am häufigsten verwendet. Diese eignen sich nur dann zur Darstellung von Geodaten, wenn der geographische Bezug nur bei der Abfrage relevant ist und die Ergebnisse unabhängig von ihrer geographischen Lage angezeigt werden sollen.

### *Digitaler Datentransfer*

Digitaler Datentransfer auf rechnerlesbaren Medien oder via Netzwerk sind Formen der Darstellung, die für den Menschen nicht lesbar sind. Daher werden sie hier nur der Vollständigkeit halber erwähnt.<sup>48</sup>

---

<sup>48</sup> Vgl. Bill, R. (1996), S. 161-172.

## Vergleich der Präsentationskonzepte

Die *interaktive Karten-Darstellung* mit Text eignet sich in ihrer Darstellungsform für Geodaten. Jedoch bedarf die interaktive Darstellung einer hohen Bandbreite, um in einer akzeptablen Geschwindigkeit ablaufen zu können.

Die *rein textuelle Darstellung* der Suchergebnisse mit kurzer Zusammenfassung, wie sie bei heutigen Suchmaschinen üblich ist, stellt die geographischen Topologien unzureichend dar. Wenn der geographische Bezug jedoch nur für die Vorauswahl der Suchergebnisse gebraucht wird und die Lage nicht als Sortierungskriterium herangezogen wird, ist die textuelle Darstellung wegen ihres vergleichsweise geringen Datentransfers der Karten-Darstellung vorzuziehen.

Die textuelle Darstellung der Suchergebnisse kann durch eine Karte ergänzt werden. So wird die übertragungsintensive Darstellung nur dann genutzt, wenn die Vorteile einer graphischen Visualisierung benötigt werden.

### 4.7 Offene Fragen

In den vorangegangenen Kapiteln wurden die Schritte, die zur Realisierung einer geographischen Suche notwendig sind, vertieft und unterschiedliche Lösungsansätze besprochen (s. Abb. 4.1).

Die Informationserhebung setzt sich aus drei Teilschritten zusammen: der *Informationserhebung* auf Web-Ebene, die sich mit der Identifizierung des geographischen Bezugs einer Web-Site beschäftigt, der *Repräsentation* des Bezugs im Web beispielsweise auf einer Homepage und der *Informationsakquisition* der Suchmaschine, die sich damit beschäftigt, die im Web veröffentlichten geographischen Informationen in der Informationsbasis der Suchmaschine strukturiert abzuspeichern. Hier existieren verschiedene, oben vorgestellte Verfahren, die in der folgenden Tabelle zusammengefasst und bewertet werden.

	Informations- erhebung und Repräsentation im Web				Informationserhebung der Suchmaschine					Bewertung	
	skalierbar	gleiches Verständnis von Geo-Bezug	benutzerfreundlich	schnell durchführbar	effizient	skalierbar	Aktualisierungsaufw.	korrekte Datenanalyse	Datenformate		Unterstützung versch. wiederverwendbar
ISO-19115-Registrierungsdienst	+	+	-	-	+	-	+	+	+	-	<b>2</b>
geo.xml - vom Inhalt getrennte RDF-Datei	+	+	+	+	+	+	+	+	+	+	<b>10</b>
HTML-Datei mit integriertem RDF	+	+	+	+	+	+	+	+	+	+	<b>10</b>
HTML-Datei mit META-Tags	+	+	+	+	+	+	+	-	+	+	<b>8</b>
HTML-Datei mit Geoinformation im Klartext	+	-	+	+	+	+	+	-	+	-	<b>4</b>
HTML-Datei mit Geoinformation im address-Tag	+	-	+	+	+	+	+	-	+	-	<b>4</b>
Web-Server-Standort	+	-	+	+	+	+	+	-	x	-	<b>3</b>

-- schlecht - 1 Punkt  
 + gut +1 Punkt  
 x kann nicht bewertet werden

Abb. 4.15: Vergleich und Bewertung der Konzepte zur Informationserhebung

In der Abb. 4.15 ist zu erkennen, dass die Zurückverfolgung der IP-Adresse den Anforderungen nur sehr schlecht genügt und daher für eine Informationserhebung dieser Art nicht geeignet ist.

Mit dem Registrierungsdienst auf der einen Seite und den Geoinformationen in einer HTML-Datei auf der anderen Seite gibt es Verfahren sowohl für sehr stark als auch schwach strukturierte Daten. Jedoch haben sie konzeptionelle Nachteile: Stark strukturierte Daten können über einen Registrierungsdienst erfasst werden, der sich inhaltlich an ISO-Standards orientiert. Diese Art der Informationserhebung verlangt vom Registrierenden professionelle Kenntnisse des verwendeten Standards. Wird ein Registrierungsdienst mit einem Redaktionsteam kombiniert, kann dieses Verfahren nicht für das ganze Internet skaliert werden. Die Indizierung von Web-Sites auf der Basis von Kontaktadressen ist zu ungenau und fehleranfällig.

Es wird daher ein Konzept benötigt, das die genaue, aber flexible Angabe von Geobezügen ermöglicht und zugleich die Informationen dezentral erfasst und somit "unbegrenzt" skalierbar ist. Die hier vorgestellten Methoden der Codierung von Metadaten in RDF oder in META-Tags entsprechen den Anforderungen am besten. Im folgenden Kapitel werden diese Verfahren näher untersucht und technische Lösungsansätze gemacht.

## 5 Technischer Ansatz

In diesem Kapitel werden für die Informationserhebung drei verschiedene technische Lösungsmöglichkeiten entwickelt. Um den geographischen Bezug einer Web-Seite in dieser zu speichern, werden Metadaten verwendet.

### 5.1 META-Tag in HTML

Als eine der verschiedenen Methoden, Metadaten über HTML-Dokumente anzugeben, wird in der HTML-Spezifikation 4.01 das META-Tag beschrieben. Dieser Ausdruck, um eine Eigenschaft von Dokumenten zu beschreiben, ist vielen HTML-Autoren vertraut und durch seine kurze und klare Schreibweise sehr einfach zu verwenden. META-Tags können in HTML auf unterschiedliche Weise implementiert werden:

- 1) Die Eigenschaft und der dazugehörige Wert werden deklariert, z. B.
  - innerhalb des Dokuments durch ein META-Tag.
  - außerhalb eines Dokuments durch Verbinden der Metadaten mit dem HTML-Dokument über das `LINK` Element.
- 2) Es wird getrennt von der HTML-Datei ein Profile mit seinen gültigen Werten definiert. Innerhalb des `HEAD` Elements wird dann durch das `profile` Attribut die URI festgelegt, unter der das Profile zu finden ist.<sup>49</sup>

Das hier gegebene Beispiel wird durch ein META-Tag innerhalb eines HTML-Dokuments realisiert, obwohl auch die Verwendung einer anderen Syntax möglich ist. Es besteht aus dem Attribut `name`, der den Namen der zu deklarierenden Eigenschaft angibt und dem Attribut `content`, in dem der Inhalt der Eigenschaft gespeichert wird.

```
<!ELEMENT META - O EMPTY -- generic metainformation -->
<!ATTLIST META
  name      NAME      #IMPLIED  -- metainformation name --
  content   CDATA     #REQUIRED -- associated information --
>
```

Abb. 5.1: Das META-Element

Abhängig davon, ob die Geo-Referenz in Koordinaten- oder in Adressform angegeben wird, kann aus verschiedenen Attributnamen ausgewählt werden. Um den Bezug an-

<sup>49</sup> Vgl. Ragget, D.; Hors, A.; Jacobs, I. (1999), Kap. 7.4.4.

hand von Koordinaten zu implementieren, wird der Attributname `geo.position` verwendet. Grundsätzlich steht die Wahl des Koordinatensystems frei, hier werden die Koordinaten jedoch in Anlehnung an den Vorschlag von Andrew Daviel in östlicher geographischer Länge und nördlicher geographischer Breite codiert. Das META-Tag `geo.position` sieht wie folgt aus:

```
<META NAME="geo.position" CONTENT="latitude; longitude">
```

Abb. 5.2: META-Tag "geo.position"

Andrew Daviel und Felix Kaegi haben eine Suchmaschine<sup>50</sup> auf ähnlicher Basis entwickelt. Das folgende Beispiel ist leicht modifiziert von der Web-Site des London-MTB Wallpaper, die auf dieser Suchmaschine registriert ist, entnommen. Sie bezieht sich auf einen Ort in der Nähe von London, Great Britain, mit den Koordinaten 51,7° N und 0,4° O.

```
<HTML>
  <HEAD profile="http://geotags.com/geo">
    <TITLE> ... </TITLE>
    <META name="geo.position" content="51.7;0.4">
  </HEAD>
  <BODY>
    ...
  </BODY>
</HTML>51
```

Abb. 5.3: META-Tag-Codierung durch Koordinaten

Wenn der geographische Bezug in textueller Form als Adresse implementiert wird, stehen dem Autor fünf Geo-Attribute zur Verfügung:

- `geo.address` beinhaltet den Straßennamen,
- `geo.no` die Hausnummer,
- `geo.postal` die Postleitzahl,
- `geo.city` bestimmt die Stadt,
- `geo.region` legt den Staat mit seinen Unterregionen fest.

<sup>50</sup> Die Suchmaschine GeoSearch ([www.geotags.com](http://www.geotags.com)) basiert auf drei geographischen Meta-Tags: Das Meta-Attribut `geo.position` gibt den geographischen Bezug in Längen- und Breitengraden, `geo.region` das Land mit seinen Unterteilungen (z.B. Bundesländer) und `geo.placename` den Ort an.

<sup>51</sup> Vgl. Source Code der Web-Site <http://www.london-mtb.com/wallpaper/>.

Das Format des letzten Attributs orientiert sich am internationalen Standard zur Darstellung von Staaten und Regionen, dem Standard ISO-2166 "Representation of Names of Countries and their subdivision".<sup>52</sup>

```
<HTML>
  <HEAD profile="http://myprofile.com/address">
    <TITLE> ... </TITLE>
    <META name="geo.address" content="Philipp Street">
    <META name="geo.no" content="584">
    <META name="geo.postal" content="E14J8G">
    <META name="geo.city" content="London">
    <META name="geo.region" content="GB-LND">
  </HEAD>
  <BODY>
    ...
  </BODY>
</HTML>
```

Abb. 5.4: META-Tag-Codierung durch eine Adresse

Neben diesem sehr einfachen Lösungsansatz gibt es noch viele andere Möglichkeiten der Spezifizierung von Metadaten. Im folgenden Kapitel wird ein sehr viel umfangreicherer Ansatz basierend auf dem Resource Description Framework vorgestellt.

## 5.2 Resource Description Framework (RDF)

Das W3C Resource Description Framework (RDF) wurde 1999 eine W3C Empfehlung für Web-Site-Autoren, um maschinenlesbare Metadaten über HTML-Dokumente oder andere Ressourcen zu spezifizieren.<sup>53</sup> RDF ist eine Basis, um Metadaten zu verarbeiten. Es stellt ein standardisiertes Datenformat zum Austausch von maschineninterpretierbaren Informationen im Web zur Verfügung und wird beispielsweise zur Verbesserung von Suchfunktionalitäten der Suchmaschinen verwendet.<sup>54</sup>

Das RDF Data Model ist ein abstraktes, konzeptionelles Framework um Metadaten zu definieren und zu nutzen. Als syntax-neutrale Darstellung benötigt das RDF Data Model

<sup>52</sup> Vgl. Daviel, A.; Kaegi, F. (2001).

<sup>53</sup> Vgl. Ragget, D.; Le Hors, A.; Jacobs, I. (1999), S. 64.

<sup>54</sup> Vgl. Lassila, O.; Swick, R. (1999), Kapitel 1.

somit eine konkrete Syntax, um die Daten tatsächlich zu erstellen und austauschen zu können.<sup>55</sup> Hier wird XML als eine der möglichen Syntaxsprachen verwendet.

### 5.2.1 RDF Data Model

Das RDF Data Model besteht aus drei Objekttypen:

- Jedes Objekt, das in RDF beschrieben wird, wird als *Ressource* bezeichnet. Eine Ressource kann im hier verwendeten Kontext also beispielsweise eine Web-Site wie [www.w3.org](http://www.w3.org) sein, aber auch eine einzelne HTML-Seite oder ein Abschnitt innerhalb einer Web-Site, der durch einen Anker, z.B. <http://www.w3.org/RDF/#specs>, gekennzeichnet wurde. Jede Ressource wird durch einen Uniform Resource Identifier (URI) eindeutig identifiziert. Prinzipiell können auch Objekte, die nicht über das Web erreichbar sind, als Ressourcen in RDF beschrieben werden. Bücher können durch ihre ISBN wie Web-Sites eindeutig identifiziert werden.<sup>56</sup>
- Eine *Eigenschaft* beschreibt ein spezifisches Charakteristikum oder ein Attribut, das zu einer Ressource gehört. Für jede Eigenschaft werden erlaubte Werte definiert.<sup>57</sup>
- Als *Statement* wird die Kombination einer spezifischen Ressource mit einer Eigenschaft und ihrem Wert bezeichnet. Die Ressource ist das "Subject", die Eigenschaft stellt das "Predicate" und der Wert der Eigenschaft das "Object" des Statements dar.<sup>58</sup>

### 5.2.2 Realisierung

Der einfache Satz "51,7° is the location, the web site <http://www.london-mtb.com> refers to." wird als RDF-Statement wie folgt dargestellt:

Subject (Ressource)	<a href="http://www.london-mtb.com">http://www.london-mtb.com</a>
Predicate (Property)	location
Object (Value of Property)	51.7°

Abb. 5.5: Beispiel für ein Statement

Dieses Statement kann in ein RDF Data Model überführt werden. In folgendem Diagramm stellen die Ovale die Ressourcen dar. Die Eigenschaften werden an einen gerichteten Pfeil geschrieben. Der Wert der Eigenschaft wird in Rechtecke eingetragen.

<sup>55</sup> Vgl. Lassila, O.; Swick, R. (1999), Kapitel 2.2.

<sup>56</sup> Vgl. Lassila, O.; Swick, R. (1999), Kapitel 2.1.

<sup>57</sup> Vgl. Lassila, O.; Swick, R. (1999), Kapitel 2.1.

<sup>58</sup> Vgl. Lassila, O.; Swick, R. (1999), Kapitel 2.1.

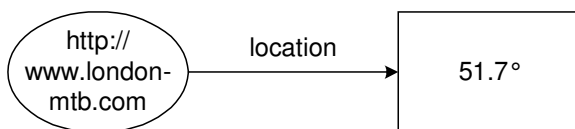


Abb. 5.6: Graphische Darstellung eines Statements

Möchte man den geographischen Bezug einer Web-Site oder einer Web-Seite codieren, benötigt man mehr Möglichkeiten der Ortsangabe. Beispielsweise kann die "location" als Punkt, Linie, Rechteck, Kreis, Polygon oder aber als Kombination mehrerer genannter Basisobjekte dargestellt werden. Des Weiteren besteht die Möglichkeit, die Ortsangabe auch textuell als Adresse zu codieren. In Abb. 5.7 wird das RDF Data Model zu den verschiedenen Darstellungsvarianten abgebildet.

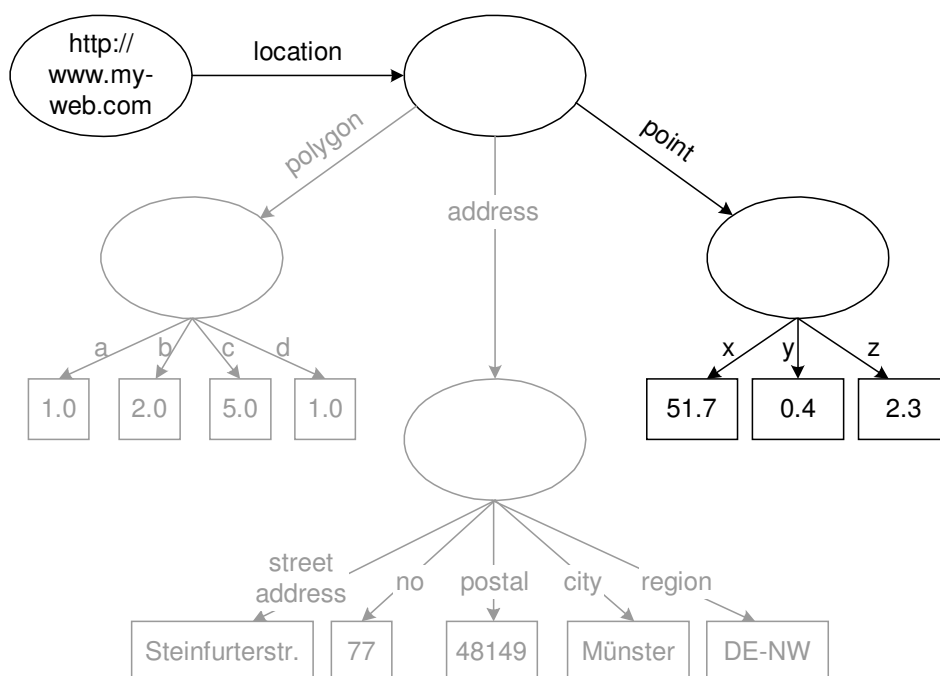


Abb. 5.7: Graphische Darstellung eines geographischen Bezugs in RDF

Wenn für eine Eigenschaft mehrere Werte existieren, wird die Eigenschaft zu einer eigenen Ressource - einer anonymen Ressource. Sie wird im RDF Data Model als leeres Oval dargestellt.

### 5.2.3 Namespace für geographische Elemente

Es müssen grundlegende Eigenschaften der Ressourcen definiert werden, damit ein geographischer Bezug in RDF kodiert werden kann. Um eine einheitliche Verwendung

der Elemente innerhalb des RDF Data Model sicherzustellen, muss ein Schema verwendet werden, in dem diese klar definiert sind.

Im Resource Description Framework besteht die Möglichkeit, externe Schemata als Namespaces einzubinden. Um geographische Objekte in XML-Syntax darzustellen, wurde die Geographic Markup Language (GML) entwickelt. Da GML 2.1.2 ein mittlerweile weit verbreiteter Standard ist, wird es hier als Namespace in RDF eingebunden, um geographische Objekte zu realisieren.

GML besteht aus drei Schemata (s. Abb. 5.8). In dem Schema Geometry.xls werden sämtliche geographischen Komponenten beschrieben.

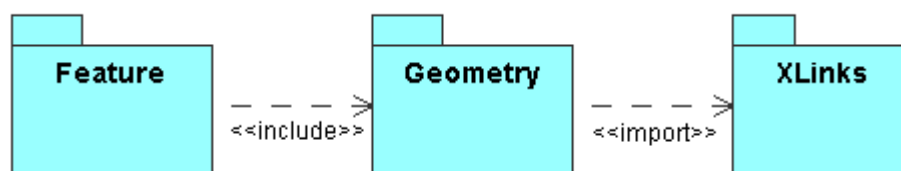


Abb. 5.8: Die drei Schemata von GML<sup>59</sup>

Mit dem Geometry Schema können verschiedene geographische Basiselemente erzeugt werden, von denen die für diesen Zweck relevanten in der nachfolgenden Tabelle aufgelistet und erklärt werden:

Geometry type	Beschreibung
Point	Stelle, Standort, Position, Mittelpunkt
LineString	Mittellinie, Rand
Polygon	Bereich
MultiPoint	eine Menge von Stellen, Standorten, Positionen, Mittelpunkten
MultiLineString	eine Menge von Mittellinien, Rändern
MultiPolygon	eine Menge von Bereichen

Abb. 5.9: Geographische Basistypen

Diese Basistypen können zur Beschreibung eines geographischen Bezugs verwendet werden. Grundlegender Bestandteil aller Basisobjekte sind die Koordinaten, die auf zwei verschiedene Arten dargestellt werden können. Hier wird nur die explizite Darstellung verwendet, aber natürlich ist auch die in der GML-Spezifikation erklärte, verkürzte Darstellung der Koordinaten möglich.

<sup>59</sup> Geography Markup Language (GML) Implementation Specification, version 2.1.2.

In GML können Objekte im dreidimensionalen Raum mit den Koordinaten X, Y und Z abgebildet werden, von denen die Angabe der Y- und Z-Koordinaten optional ist. Die Koordinaten werden als Dezimalzahlen codiert.

### Point

Ein Punkt besteht aus einem einzigen Koordinatentupel, das mindestens eine X-Koordinate und optional eine Y- und eine Z-Koordinate hat. Abb. 5.10 stellt die xml-notierte Definition im Geometry Schema durch den XML-Spy graphisch dar.

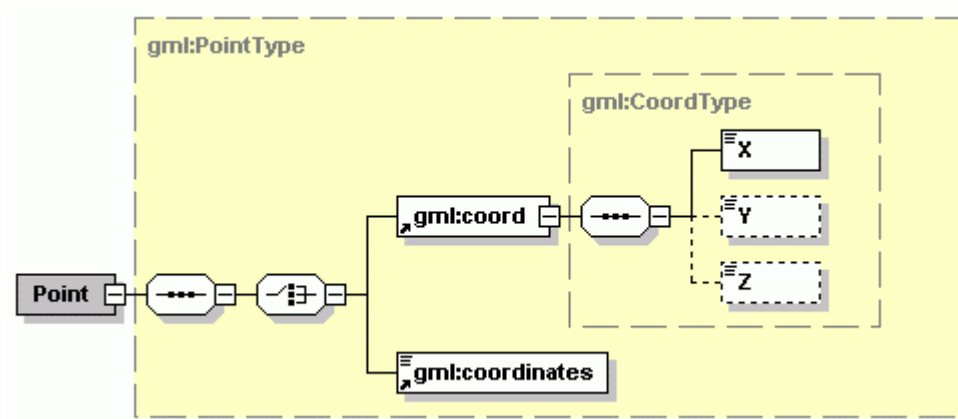


Abb. 5.10: Schema für einen Point in GML<sup>60</sup>

Um den geographischen Bezug der Web-Site des London-MTB Wallpaper im GML-Notation zu codieren, muss der folgende Code angegeben werden:

```
<Point>
  <coord><X>51.7</X><Y>0.4</Y></coord>
</Point>
```

Abb. 5.11: Beispielcode für einen Point in GML

### LineString

Zur Darstellung des LineString werden mindestens zwei Koordinaten gebraucht, die Anfangs- und Endpunkt der Line festlegen. Die Punkte auf der Linie werden durch lineare Interpolation ermittelt.

<sup>60</sup> Diese und die folgenden Abbildungen wurden mit der XML Spy 5 Professional Edition erzeugt.

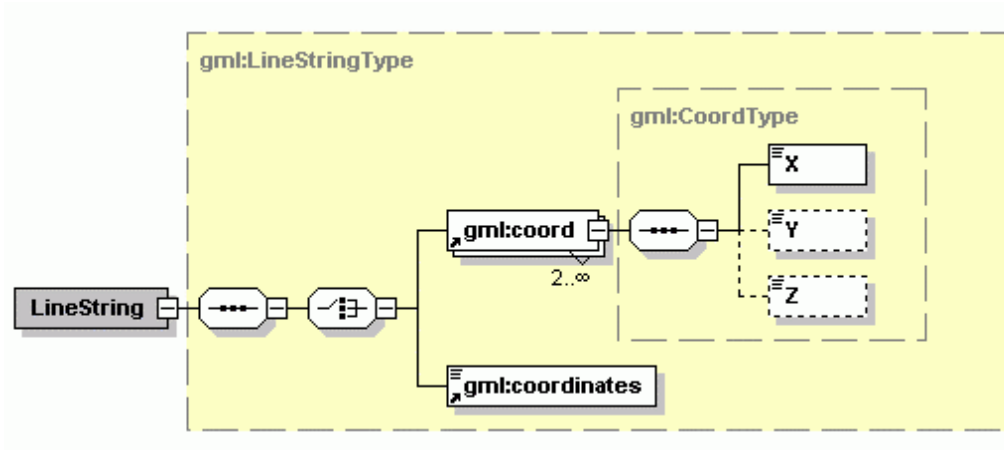


Abb. 5.12: Schema für einen LineString in GML

Falls mehrere Punkte angegeben sind, werden sie miteinander verbunden, sodass ein stückweise linearer Weg entsteht.

```
<LineString>
  <coord><X>53.2</X><Y>0.2</Y></coord>
  <coord><X>52.9</X><Y>1.7</Y></coord>
</LineString>
```

Abb. 5.13: Beispielcode für einen LineString in GML

Das Beispiel in Abb. 5.13 zeigt einen geographischen Bezug, der sich auf einen Abschnitt von Großbritanniens Küste bezieht. Eine Reisegesellschaft könnte so beispielsweise ihr Reiseangebot an dem Badestrand in England beschreiben.

### Polygon

Ein Polygon setzt sich in der Geographic Markup Language durch einen oder mehrere LinearRing-Objekte zusammen. Ein LinearRing besteht aus mindestens drei unterschiedlichen Koordinaten. Die vierte Koordinate, der Endpunkt, muss mit der ersten, dem Anfangspunkt, übereinstimmen.

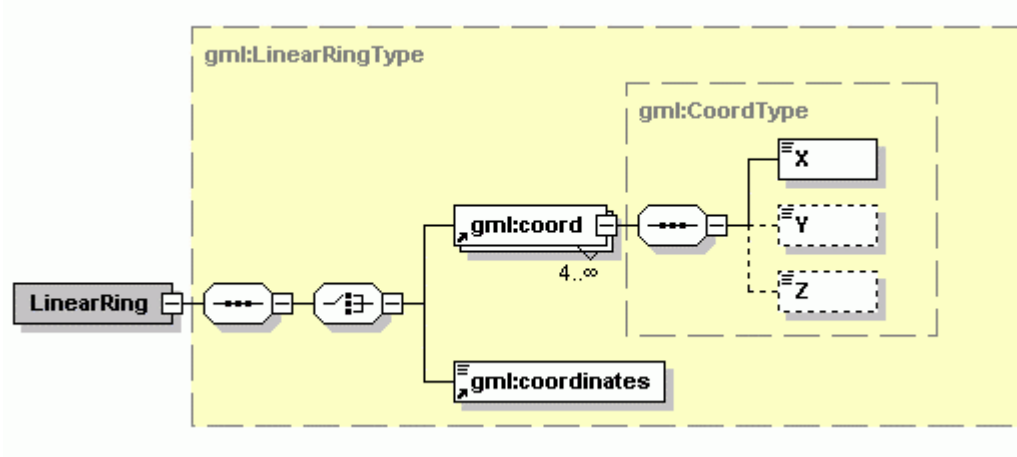


Abb. 5.14: Schema für ein Polygon in GML [1]

Ein LinearRing wird durch den Tag `<outerBoundaryIs>` als äußerer Ring definiert und stellt die Grenze des Polygons dar. Es können nun beliebig viele innere Ringe definiert werden, deren Koordinaten innerhalb des Polygons liegen. Diese markieren einen Bereich, der nicht zum Polygon gehört. Dieser Sachverhalt wird in der Abb. 5.15 graphisch dargestellt.

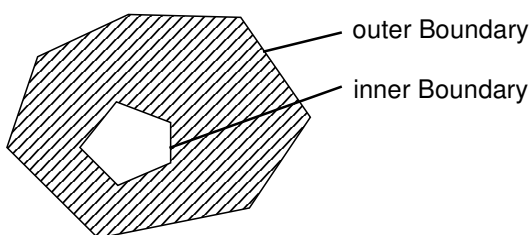


Abb. 5.15: Beispiel für ein Polygon

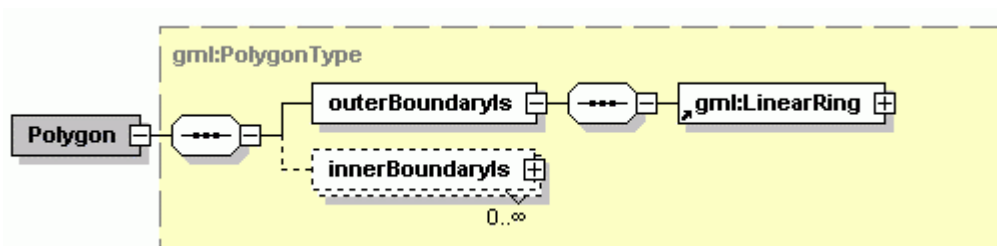


Abb. 5.16: Schema für ein Polygon in GML [2]

Polygone als geographischer Bezug können genutzt werden, um ein Kundeneinzugsgebiet festzulegen, das Einflussgebiet eines Projektes darzustellen oder auch ein Firmengelände genau zu markieren.

```
<Polygon>
  <outerBoundaryIs>
```

```

<LinearRing>
  <coord><X>40.5</X><Y>-7.7</Y></coord>
  <coord><X>42.1</X><Y>-1.8</Y></coord>
  <coord><X>37.7</X><Y>-3.3</Y></coord>
  <coord><X>37.9</X><Y>-6.3</Y></coord>
  <coord><X>40.5</X><Y>-7.7</Y></coord>
</LinearRing>
</outerBoundaryIs>
<innerBoundaryIs>
  <LinearRing>
    <coord><X>39.1</X><Y>-5.7</Y></coord>
    <coord><X>39.9</X><Y>-3.8</Y></coord>
    <coord><X>39.1</X><Y>-5.7</Y></coord>
  </LinearRing>
</innerBoundaryIs>
</Polygon>

```

Abb. 5.17: Beispielcode für ein Polygon in GML

In diesem Beispiel wurde ein Gebiet in Spanien markiert, bei dem in der Mitte ein dreieckiger Bereich ausgespart ist. Da hierbei sehr viele Koordinaten kodiert werden müssen, bietet es sich an, die verkürzte Koordinatenschreibweise von RDF zu verwenden.<sup>61</sup>

*MultiPoint, MultiLineString, MultiPolygon*

In GML kann auch eine Sammlung von geographischen Basisobjekten gespeichert werden. Hier wird stellvertretend für alle Multi-Objekte nur der MultiPoint behandelt. Die anderen Konstrukte werden nach dem gleichen Schema gebildet.

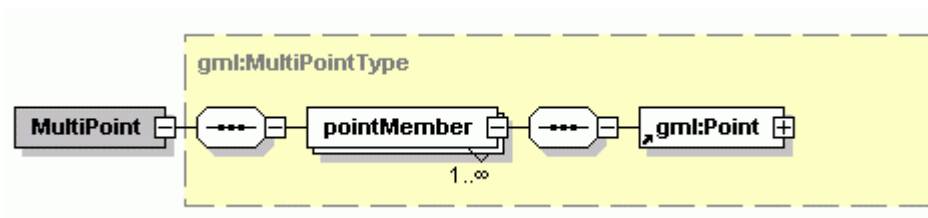


Abb. 5.18: Schema für einen MultiPoint in GML

Der MultiPoint besteht aus einer Auflistung von Punkten, die durch das Tag <pointMember> deutlich gemacht werden. Innerhalb des <pointMember> Elements wird ein Punkt nach der oben vorgestellten Weise implementiert.

<sup>61</sup> Geography Markup Language (GML) Implementation Specification, version 2.1.2, Kapl 4.3.1.

```

<MultiPoint>
  <PointMember>
    <Point>
      <coord><X>39.2</X><Y>-2.1</Y></coord>
    </Point>
  </PointMember>
  <PointMember>
    <Point>
      <coord><X>55.4</X><Y>-2.7</Y></coord>
    </Point>
  </PointMember>
</MultiPoint>

```

Abb. 5.19: Beispielcode für einen MultiPoint in GML

Der in Abb. 5.19 dargestellte geographische Bezug kann beispielsweise dazu genutzt werden, um zwei Standorte einer Firma in Spanien und Großbritannien zu speichern.

Neben der Geographic Markup Language können auch andere Namespaces wie z.B. der ISO-19115-Standard "Geographic Information - Metadata" verwendet werden. Da der ISO-19115-Standard zur Darstellung geographischer Objekte auch GML benutzt, wird die Geographic Markup Language direkt integriert.

### 5.2.4 Namespace für textuelle Elemente

Die geographische Adresse kann auch durch textuelle Adressangaben kodiert werden. In der folgenden Abbildung ist dieses Schema graphisch dargestellt. Wie bei der Implementierung durch META-Tags werden auch hier die verschiedenen Properties `streetaddress`, `no`, `postal`, `city` und `region` unterschieden.

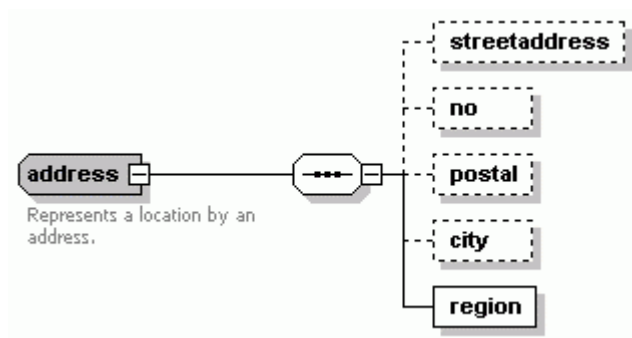


Abb. 5.20: Schema für eine Adresse [1]

Der Straßenname `streetaddress`, die Hausnummer `no`, die Postleitzahl `postal` und die Stadt `city` sind hier als optionale Elemente definiert, da so Ortsangaben von Web-Sites, die regionalen Bezug haben, durch Angabe der Region gespeichert werden können. Die Region wird wie zuvor bei den META-Tags nach dem Standard ISO-2166 "Representation of Names of Countries and their subdivision" realisiert.

```
<address>
  <streetaddress>Steinfurterstr.</streetaddress>
  <no>77</no>
  <postal>48149</postal>
  <city>Muenster</city>
  <region>DE-NW</region>
</address>
```

Abb. 5.21: Beispielcode für eine Adresse

Für diese textuelle Darstellung muss ein XML-Schema gemäß der W3C Spezifikation erstellt und unter einer URL veröffentlicht werden. Als Beispiel wird die URL <http://www.mySchema.net/adr> verwendet, unter der das Schema `address.xsd` zu finden ist. In Abb. 5.22 ist dieses Schema abgebildet.

```
<?xml version="1.0" encoding="UTF-8" ?>
<schema targetNamespace="http://www.mySchema.net/adr"
  xmlns:adr="http://www.mySchema.net/adr"
  xmlns="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified"
  version="1.00">
  <annotation>
    <appinfo>address.xsd v1.00 2002-11</appinfo>
    <documentation xml:lang="en">adr address schema.</documentation>
  </annotation>
  <complexType name="address">
    <annotation>
      <documentation>Represents a location by an address.</documentation>
    </annotation>
    <sequence>
      <element name="streetaddress" minOccurs="0" />
      <element name="no" minOccurs="0" />
      <element name="postal" minOccurs="0" />
      <element name="city" minOccurs="0" />
```

```

        <element name="region" />
    </sequence>
</complexType>
</schema>

```

Abb. 5.22: Schema für eine Adresse [2]

Die Definition dieses Namespace ist nur als ein Beispiel für eine mögliche Lösung zu sehen. Selbstverständlich kann auch eine andere Kodierung abhängig vom gewünschten Detaillierungsgrad gewählt werden. Auch die ISO-19115 "Geographic Information - Metadata" beinhaltet textuelle Ortsinformationen, da diese jedoch nicht so umfangreich sind, wurde hier für die Darstellung ein selbst definiertes Schema bevorzugt.

### 5.2.5 Lösungsvorschlag in XML-Syntax

Da ein XML-Dokument nicht ausschließlich aus RDF-Informationen bestehen muss, wird anfangs durch das Tag `<rdf:RDF>` kenntlich gemacht, wo die RDF-Informationen, die in das RDF Data Model übersetzt werden können, beginnen, und durch das Schlusstag `</rdf:RDF>`, wo sie enden. Im Anfangstag werden üblicherweise alle verwendeten Namespaces aufgeführt, obwohl diese auch erst bei Gebrauch definiert werden können. Alle Informationen über ein Dokument, d.h. die Eigenschaften mit ihren Werten, werden von den beiden Tags `<rdf:Description>` und `</rdf:Description>` eingeschlossen.

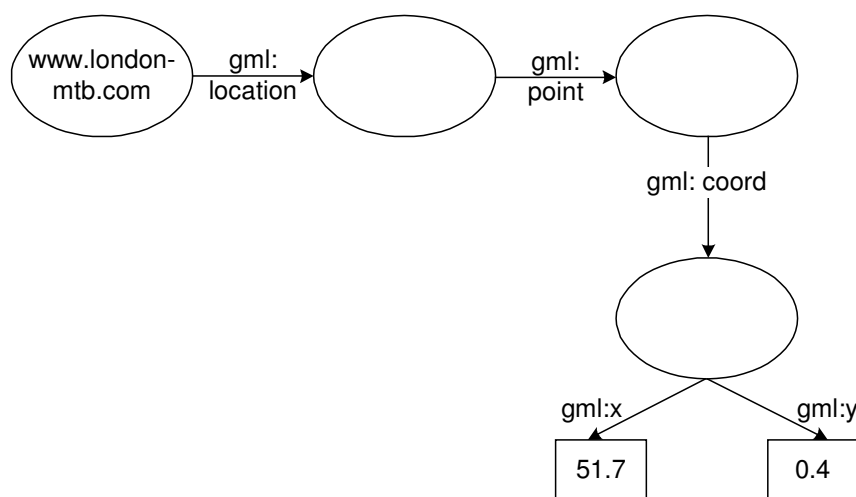


Abb. 5.23: Graphische Darstellung einer RDF-Codierung mit GML Point

Der folgende Beispielcode stellt den in der Abb. 5.23 beschriebenen geographischen Bezug der Web-Site <http://www.london-mtb.com/> dar. Der Bezug ist ein Punkt mit den Koordinaten X = 51.7 und Y = 0.4.

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:gml="http://www.opengis.net/gml#">
<rdf:Description about="http://www.london-mtb.com/">
  <gml:location rdf:parseType="Resource">
    <gml:Point rdf:parseType="Resource">
      <gml:coord rdf:parseType="Resource">
        <gml:X>51.7</gml:X>
        <gml:Y>0.4</gml:Y>
      </gml:coord>
    </gml:Point>
  </gml:location>
</rdf:Description>
</rdf:RDF>
```

Abb. 5.24: RDF-Codierung mit GML Point in XML

Als Namespaces wurden die GML-Spezifikation 2.1.2 und die RDF-Spezifikation eingebunden.

Um den Bezug durch andere geographische Objekte festzulegen, muss das Tag `<gml:location>` durch die in Kap. 5.2.4 beschriebenen Elemente ersetzt werden.

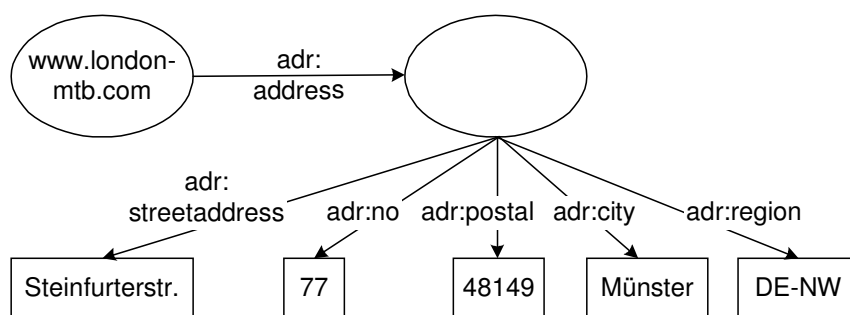


Abb. 5.25: Graphische Darstellung einer RDF-Codierung mit Adresse

Der geographische Bezug der Web-Seite `www.schwing.de/angela.html` in Form einer Adresse gemäß dem in Kap. 5.2.4 definierten Schema wird in Abb. 5.25 graphisch dargestellt und im folgenden Beispiel implementiert:

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  ...
```

```

xmlns:adr="http://www.mySchema.net/adr#">
<rdf:Description about="http://www.schwering.de/angela.html">
  <adr:address rdf:parseType="Resource">
    <adr:streetaddress>Steinfurterstr.</adr:streetaddress>
    <adr:no>77</adr:no>
    <adr:postal>48149</adr:postal>
    <adr:city>Muenster</adr:city>
    <adr:region>DE-NW</adr:region>
  </adr:address>
</rdf:Description>
</rdf:RDF>

```

Abb. 5.26: RDF-Codierung durch eine Adresse in XML

### 5.3 Resource Description Framework in HTML

Da RDF wohlgeformtes XML ist, kann der RDF-Code direkt in das <HEAD>-Tag eines HTML-Dokuments integriert werden. Ein Browser, der die HTML Empfehlung zur Fehlerbehandlung von ungültigen HTML-Dokumenten befolgt, wird jeden ungeschützten Text, d.h. alles, was außerhalb von Tags steht, anzeigen. Integriert man also den RDF-Code, so wie es in Abb. 5.27 gezeigt wird, werden die Koordinaten im Browserfenster angezeigt (s. Abb. 5.28).

```

<html>
<head>
  <rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:gml="http://www.opengis.net/gml#">
    <rdf:Description about="http://www.london-mtb.com/">
      <gml:location rdf:parseType="Resource">
        <gml:Point rdf:parseType="Resource">
          <gml:coord rdf:parseType="Resource">
            <gml:X>51.7</gml:X>
            <gml:Y>0.4</gml:Y>
          </gml:coord>
        </gml:Point>
      </gml:location>
    </rdf:Description>
  </rdf:RDF>
</head>
<body>

```

```
<P>Content of the document.</P>
</body>
</html>
```

Abb. 5.27: RDF-Codierung mit Koordinaten in HTML [1]

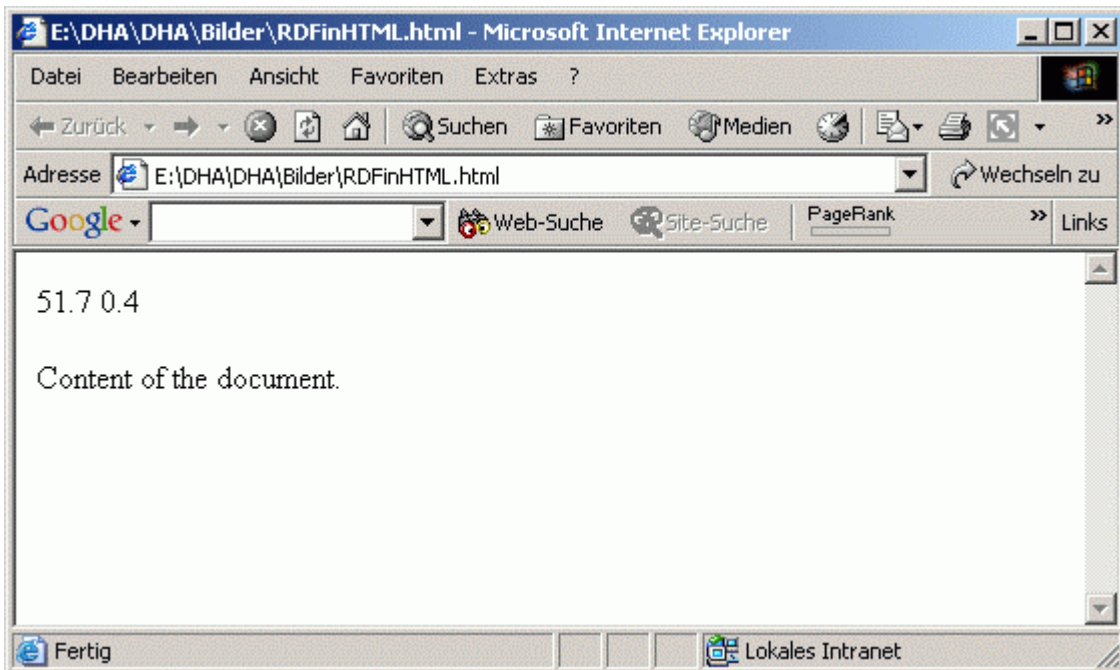


Abb. 5.28: Interpretation des RDF-Codes in einem Browser

Dieses Problem kann durch die gekürzte Syntax von RDF, die keinen ungeschützten Text enthält, behoben werden. Um den GML-Namespace in RDF zu verwenden, muss auch dieser in gekürzter Schreibweise dargestellt werden.

```
<html>
<head>
  <rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:gml="http://www.opengis.net/gml#">
    <rdf:Description about="http://www.london-mtb.com/">
      <gml:location rdf:parseType="Resource">
        <gml:Point rdf:parseType="Resource">
          <gml:coord
            gml:X="51.7"
            gml:Y="0.4"/>
        </gml:Point>
      </gml:location>
    </rdf:Description>
```

```

</rdf:RDF>
</head>
<body>
<P>Content of the document.</P>
</body>
</html>

```

Abb. 5.29: RDF-Codierung mit Koordinaten in HTML [2]

Abb. 5.29 zeigt einen Code, bei dem Koordinaten zur GML-Point-Darstellung innerhalb HTML gespeichert werden. Browser interpretieren nur den Inhalt des <Body>.

### 5.4 Diskussion

Nachdem verschiedene Ansätze beschrieben wurden, durch die "traditionelle" Web-Sites mit geographischen Informationen erweitert werden können, werden sie nun nach Aufwand und Nutzen bewertet.

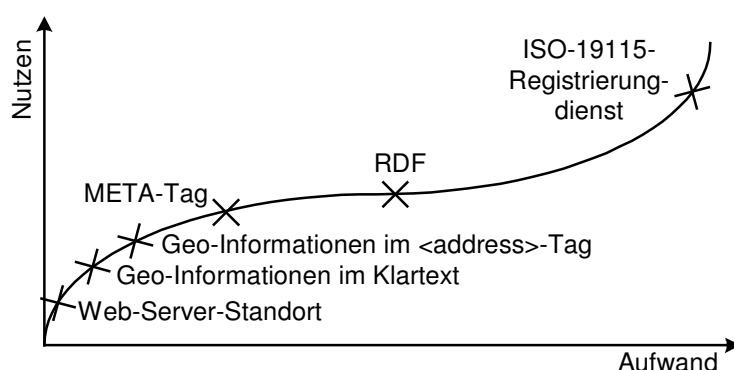


Abb. 5.30: Aufwand-Nutzen-Relation

Den geographischen Bezug aus dem *Standort des Web-Servers* zu ermitteln ist sehr wenig Aufwand, da die Daten zu jeder Web-Site vorhanden und eindeutig interpretierbar sind. Da der Bezug aber nur auf Landes- oder maximal auf regionaler Ebene festzustellen ist, hat diese Methode einen sehr geringen Nutzen.

Auf vielen Web-Sites existieren bereits *Geoinformationen, die im Klartext oder innerhalb von <address> Tags* gespeichert sind. Die Intentionen dieser Codierung sind meist aber sehr verschieden und decken sich nicht immer mit dem des geographischen Bezugs. Des Weiteren sind solche Informationen aufgrund ihrer fehlenden Struktur nur schlecht zu identifizieren. Der Aufwand ist also größer als bei der Verwendung des Web-Server-Standorts, aber durch die detailliertere geographische Genauigkeit ist der Nutzen auch höher.

Die Verwendung von *META-Tags* ist sehr einfach und schnell zu erlernen, da sie den meisten HTML-Autoren bereits bekannt sind. Der geographische Bezug wird explizit auf die Web-Site codiert. Negativ auf die geographische Genauigkeit wirkt sich jedoch aus, dass mit META-Tags nur die primitivste Form des Geo-Bezugs implementiert werden kann - der Punkt. Komplexere Gebilde wie Linien und Polygone unterstützt das hier definierte META-Tag nicht. Des Weiteren sind META-Tags beschränkt auf HTML-Dokumente. Da viele Dateien im Web in anderen Formaten vorhanden sind, kann dieses Konzept nur den auf HTML beschränkten Teilbereich des Webs mit den zusätzlichen Metadaten ausstatten.

Metadaten durch das *Resource Description Framework* zu implementieren ist gegenüber META-Tags eine sehr leistungsfähige Variante. Durch Einbinden von Namespaces können komplexe geographische Objekte beschrieben werden, die jedoch noch keine detaillierten Kenntnisse geographischer Terminologie erfordern. Dieses Konzept kann wegen seiner offenen Architektur flexibel erweitert werden, beispielsweise durch Ideen des Semantic Web und Ontologien.

Wird der in RDF implementierte Bezug in einer eigenen XML-Datei gespeichert, in der durch eine URL die Bezugsdatei festgelegt wird, so können sämtliche Dateiformate unterstützt werden. Da die Metadaten von den beschriebenen Daten einzeln gespeichert werden, ist der Verwaltungsaufwand relativ hoch. Konsistente Daten zu sichern wird durch die getrennte Datenhaltung erschwert.

Es gibt aber auch Methoden, RDF in eine Datei zu integrieren. Als wohlgeformtes XML kann RDF mit einer abgekürzten Schreibweise in HTML integriert werden. Werden die Bestrebungen durchgesetzt, XML als zusätzliches Dateiformat in Microsoft Office 11 (interner Name) zu etablieren, besteht auch die Möglichkeit, RDF in Word, Exel, PowerPoint und Access einzufügen.<sup>62</sup> Es existiert bereits ein Ansatz von W3C, wie RDF Metadaten in Bildern im .jpeg Format gespeichert werden können.<sup>63</sup>

Ein *ISO-19115-Registrierungsdienst* orientiert sich an bestehenden Standards und ist sehr stark strukturiert. Die Angaben sind sehr umfangreich und erfordern daher professionelle Kenntnisse von dem Standard und geographischer Terminologie. Daher ist dieses Konzept mit nicht akzeptablem Aufwand verbunden und für breite Internetlösungen nicht geeignet.

---

<sup>62</sup> Brors, D. (2002), S. 56.

Schüler, P. (2002), S. 49.

<sup>63</sup> Lafon, Y.; Bos, B. (2002).

## 6 Zusammenfassung und Ausblick

Effiziente Suchstrategien sind die Voraussetzung für eine erfolgreiche Informationssuche im World Wide Web. Geographische Daten im Web ermöglichen eine raumbezogene Strukturierung der Informationen und somit die Entwicklung einer neuartigen Suchstrategie - der geographischen Suche.

Die geographische Suchmaschine durchsucht das Web kontinuierlich nach Geoinformationen, d.h. nach geographischen Bezügen von Web-Sites, welche dann als Suchkriterien vom Nutzer verwendet werden können. Zur Analyse kann die Suchmaschine auf geographische Informationen zurückgreifen, die bereits auf der Web-Site vorhanden sind, oder aus IP-Nummer oder Domain-Name selbstständig auf den geographischen Bezug schließen. Da diese inhaltlichen oder protokollspezifischen Daten einer Web-Site keine verbindlichen Rückschlüsse darauf gestatten, muss der geographische Bezug des Inhalts einer Web-Site explizit codiert werden. Nur so können Genauigkeit und Qualität der von der Suchmaschine identifizierten geographischen Bezüge erhöht werden. Dazu wurden drei Ansätze entwickelt, die basierend auf Metadaten für die Informationserhebung eingesetzt werden können.

Im ersten Ansatz wird der geographische Bezug in META-Tags, die in das HTML-Dokument integriert sind, implementiert. Sowohl für die Nutzung von Koordinaten als auch von Adressen sind solche Tags für die Darstellung des geographischen Bezugs entwickelt worden. Da Bestandteile der HTML-Spezifikation genutzt werden, ist diese einfache und unkomplizierte Form der Codierung den meisten HTML-Autoren bereits bekannt und auch für unerfahrene Web-Site-Programmierer schnell durchschaubar. Andererseits ist durch die Verwendung von META-Tags der Einsatz auf HTML-Dokumente beschränkt. Ein Großteil des Webs, das aus anderen Formaten besteht, kann somit nicht erfasst werden. Durch die simple Struktur der META-Tags ist die Ausdrucksmöglichkeit, den geographischen Bezug zu beschreiben, auf einfache geographische Objekte begrenzt.

Bei den anderen beiden Konzepten können dagegen komplexe geographische Gebilde verschiedenen Dokumenten zugeordnet werden. In dem einen Ansatz wird die reine XML-Darstellung des Resource Description Framework (RDF) verwendet, im anderen Ansatz wird das RDF in HTML integriert. Um verschiedene geographische Objekte darstellen zu können, wird empfohlen, die Geographic Markup Language (GML) als Namespace in RDF einzubinden. Ein geographischer Ort kann auch durch Angabe einer Adresse eindeutig angegeben werden. Hierfür wird ein eigenes Schema in XML entwickelt, das die erforderlichen Bestandteile einer Adresse festlegt. Im dritten Ansatz, der

Codierung des Resource Description Framework innerhalb HTML, wird eine abgekürzte Schreibweise für GML vorgeschlagen.

Verglichen mit dem ersten Ansatz ist die Implementierung der Metadaten durch das Resource Description Framework ein sehr leistungsfähiges und offenes Konzept. Über Namespaces kann der Ansatz variabel durch andere Metadaten-Standards erweitert werden, wie beispielsweise durch die kürzlich für RDF entwickelte vereinfachte Syntax des Dublin Core Metadata Element Set. Durch die Offenheit der Architektur besteht die Möglichkeit, das Konzept in Zukunft flexibel im Sinne des Semantic Web und von Ontologien zu erweitern.

Wird der in RDF implementierte Bezug in einer eigenen Datei gespeichert, können sämtliche Dateiformate unterstützt werden. Die Bestrebungen, RDF in bestehende Formate einzubeziehen, werden intensiver: Neben der Integration in HTML-Dokumente gibt es Ansätze, RDF-Metadaten über Bilder in das .jpeg Format einzufügen. Wird XML als zusätzliches Dateiformat in Microsoft Office eingeführt, können in Zukunft sämtliche Office-Dokumente durch Metadaten beschrieben werden. Die Kombination des reinen XML-orientierten RDF mit dem in HTML integrierten RDF wird als optimaler Lösungsansatz einer geographischen Suchmaschine vorgeschlagen.

## Literaturverzeichnis

- Bartelme, N.: Geoinformatik - Modelle, Strukturen, Funktionen. Springer Berlin u.a., 2000.
- Bill, R.; Zehner, M.: Lexikon der Geoinformatik. Herbert Wichmann Verlag Heidelberg, 2001.
- Bollmann, J.; Gunther, W.: Lexikon der Kartographie und Geomatik, Band 1 "A bis Karti". Spektrum Akademischer Verlag Heidelberg u.a., 2001.
- Brors, D.: Post-XP-Ära - Erster Eindruck von Microsoft Office 11. In C'T (2002) 25, S. 56, <http://www.heise.de/ct/02/25/056/>, [02.12.2002].
- Chang, E.: Future Search Engines. Of Robots, Spiders and Worms - The Search Begins. 1996, [http://www.doc.ic.ac.uk/~nd/surprise\\_96/journal/vol1/ec5/article1.html](http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/ec5/article1.html), [14.11.2002].
- Cox, S.; Cuthbert, A.; Lake, R.; Martell, R.: OpenGIS® Geography Markup Language (GML) Implementation Specification, version 2.1.2. 17. September 2002, <http://www.opengis.net/gml/02-069/GML2-12.html>, [17.11.2002].
- Daviel, A.; Kaegi, F.: Geographic extensions for HTTP transactions. April 2001, <http://geotags.com/geo/draft-daviel-http-geo-header-03.txt>, [2002-11-19].
- Daviel, A.; Kaegi, F.: Geographic registration of HTML documents. April 2001, <http://geotags.com/geo/draft-daviel-html-geo-tag-05.txt>, [2002-11-19].
- Davis, C.: DNS-LOC: Geo-enabling the Domain Name System. 2001, <http://www.ckdhr.com/dns-loc/> [15.12.2002].
- Egenhofer, M.: Why not SQL? In: International Journal of Geographical Information Systems. 2 (1992) 6, S. 71-85.
- Findeisen, D.: Datenstrukturen und Abfragesprachen für raumbezogene Informationen. In: Schriftenreihe des Instituts für Kartographie und Topographie der Rheinischen Friedrich-Wilhelms-Universität Bonn. Hrsg.: Morgenstern, D.; Kupfer, G., Kirschbaum Verlag Bonn (1990) 19.
- Günther, O.: Environmental Information Systems. Springer Berlin u.a., 1998.
- Kurbel, K.: Entwicklung und Einsatz von Expertensystemen - Eine anwendungsorientierte Einführung in wissensbasierte Systeme. Springer Verlag Berlin u.a., 1992.
- Lafon, Y.; Bos, B.: Describing and retrieving photos using RDF an HTTP. W3C Note, 19 April 2002, <http://www.w3.org/TR/photo-rdf/>, [01.12.2002].

- Lassila, O.; Swick, R.: Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation, 22 February 1999, <http://www.w3.org/TR/REC-rdf-syntax>, [20.10.2002].
- Manjoo, F.: Google's Search for Winner Ends. 31.05.2002, <http://www.wired.com/news/infostructure/0,1377,52897,00.html>, [10.09.2002].
- Mertens, P, u.a.: Lexikon der Wirtschaftsinformatik. Springer Berlin u.a., 2001.
- O.A.: 2002 Google Programming Contest Winner. <http://www.google.com/programming-contest/winner.html>, [10.09.2002].
- O.A.: Geometry Schema. Open GIS Consortium, 16.02.2001, <http://www.opengis.net/namespace/gml/core/geometry.xsd>, [02.12.2002].
- O.A.: World Wide Web in 7 Points. World Wide Web Consortium, <http://www.w3.org/Consortium/Points/>, [14.11.2002].
- Ragget, D.; Le Hors, A.; Jacobs, I.: HTML 4.01 Specification. W3C Recommendation, 24 December 1999, <http://www.w3.org/TR/html401/>, [07.11.2002].
- Röscheisen, E.: Spurensuche. In iX (2002) 8, S. 90, <http://www.heise.de/ix/artikel/2002/08/090/>, [15.12.2002].
- Röscheisen, E.: Von IP-Adressen zu Ortinformationen. In iX (2002) 8, S. 90, <http://www.heise.de/ix/artikel/2002/08/090/02.html>, [15.12.2002].
- Schneider, B.: Vorlesung Künstliche Intelligenz in der Betrieblichen Anwendung Kapitel 2.2.5. <http://www.wi.uni-muenster.de/wi/lehre/wbs/ws01-02/kibakapitel225.pdf>, [10.10.2002].
- Schüler, P.: Hickhack um Office XML - Ein Arbeitskreis soll offenen Standard schaffen. In: C'T (2002) 25, S. 49.
- Schumann, H.; Müller, W.: Visualisierung – Grundlagen und allgemeine Methoden. Springer Heidelberg, 2000.
- Seidenberg, U.: Ist Information als eigenständiger Produktionsfaktor aufzufassen? Siegen, 1998, Arbeitspapier: <http://www.uni-siegen.de/dept/fb05/prod/pfaktinf.pdf>, [13.12.2002].
- Steinberg, S.: Seek an Ye Shall Find (Maybe). 1996, [http://hotwired.lycos.com/collections/web\\_development/4.05\\_search\\_pr.html](http://hotwired.lycos.com/collections/web_development/4.05_search_pr.html), [14.11.2002].
- Streit, U.: Vorlesung Einführung in die Geoinformatik, Version 3.1. <http://castafiore.uni-muenster.de/vorlesungen/geoinformatik/>, [23.09.2002].

Tanenbaum, A.: Computernetzwerke. Pearson Studium München, 2000.

Teubner, R.: Integrierte Organisations- und Informationssystemgestaltung: Methoden für das Organization und Information Engineering. Univ.Diss. Münster, 1997.

Teuteberg, F.: Effektives Suchen im World Wide Web: Suchdienste und Suchmethoden. Wirtschaftsinformatik 39 (1997) 4, S. 373-383.

## Anhang

### A Ortsbezug durch IP-Adressen

```
uni-muenster.de
Request: uni-muenster.de

domain:      uni-muenster.de
descr:      Westfaelische Wilhelms-Universitaet
descr:      Zentrum fuer Informationsverarbeitung
descr:      Roentgenstrasse 9-13
descr:      48149 Muenster
descr:      Germany
nserver:    comsrv02.uni-muenster.de 128.176.0.12
nserver:    comsrv03.uni-muenster.de 128.176.0.13
nserver:    noc.hrz.uni-bielefeld.de
nserver:    ws-leil.win-ip.dfn.de
status:     connect
changed:    20021105 120608
source:     DENIC

[admin-c]
Type:       PERSON
Name:       Wilhelm Held
Address:    Westfaelische Wilhelms-Universitaet
Address:    Zentrum fuer Informationsverarbeitung
Address:    Roentgenstrasse 9-13
City:      Muenster
Pcode:     48149
Country:    DE
Changed:    20021105 114516
Source:     DENIC

[tech-c][zone-c]
Type:       PERSON
Name:       Georg Richter
Address:    Westfaelische Wilhelms-Universitaet
Address:    Zentrum fuer Informationsverarbeitung
Address:    Roentgenstrasse 9-13
```

```
City:      Muenster
Pcode:     48149
Country:   DE
Phone:     +49 251 83 31611
Fax:       +49 251 83 31653
Email:     network@uni-muenster.de
Changed:   20021105 094440
Source:    DENIC
```

**Anh. A.1: Informationen eines WWW-Whois-Dienstes<sup>64</sup>**

```
OrgName:   Westfaelische Wilhelms - Universitaet
OrgID:     WWU-1

NetRange:  128.176.0.0 - 128.176.255.255
CIDR:      128.176.0.0/16
NetName:   DMSWWU-ETHER
NetHandle: NET-128-176-0-0-1
Parent:    NET-128-0-0-0-0
NetType:   Direct Assignment
NameServer: COMSRV02.UNI-MUENSTER.DE
NameServer: COMSRV03.UNI-MUENSTER.DE
NameServer: NOC.HRZ.UNI-BIELEFELD.DE
NameServer: WS-LEI1.WIN-IP.DFN.DE
Comment:
RegDate:   1986-11-13
Updated:   1999-02-16

TechHandle: GR26-ARIN
TechName:  Richter, Georg
TechPhone: +49 251 83 31611
TechEmail: network@uni-muenster.de

# ARIN Whois database, last updated 2002-12-14 20:00
# Enter ? for additional hints on searching ARIN's Whois database.
```

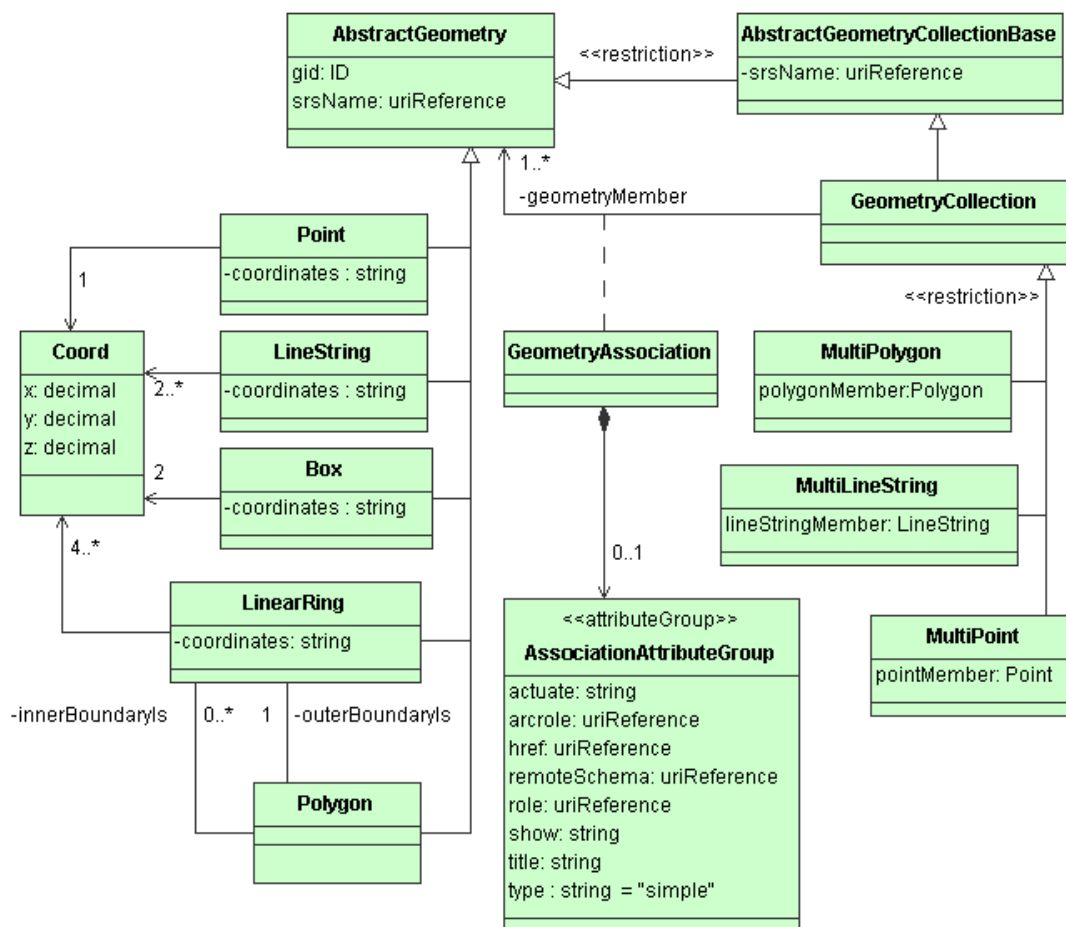
**Anh. A.2: Informationen eines IP-Whois-Dienstes<sup>65</sup>**

---

<sup>64</sup> Diese Informationen sind von dem Dienst ALLWHOIS entnommen, <http://www.allwhois.com>.

<sup>65</sup> Diese Informationen sind von dem Dienst ip-telligence entnommen, <http://ailab.de>.

## B Geometry Schema der Geographic Markup Language



Anh. B.1: UML-Darstellung für das Geometry Schema<sup>66</sup>

```

<element name="coord" type="gml:CoordType" />
= <complexType name="CoordType">
= <annotation>
= <documentation>Represents a coordinate tuple in one, two,
= or three dimensions.</documentation>
= </annotation>
= <sequence>
= <element name="X" type="decimal" />
= <element name="Y" type="decimal" minOccurs="0" />
= <element name="Z" type="decimal" minOccurs="0" />
= </sequence>
= </complexType>

```

Anh. B.2: Geometry Schema für die Coordinates<sup>67</sup>

<sup>66</sup> Geography Markup Language (GML) Implementation Specification, version 2.1.2.

<sup>67</sup> Vgl. Geometry Schema.

```

<element name="Point" type="gml:PointType" substitution-
  Group="gml:_Geometry" />

= <complexType name="PointType">
  = <annotation>
    <documentation>A Point is defined by a single coordinate tu-
      ple.</documentation>
  </annotation>
  = <complexContent>
    = <extension base="gml:AbstractGeometryType">
      = <sequence>
        = <choice>
          <element ref="gml:coord" />
          <element ref="gml:coordinates" />
        </choice>
      </sequence>
    </extension>
  </complexContent>
</complexType>

```

Anh. B.3: Geometry Schema für den Point<sup>68</sup>

```

<element name="LineString" type="gml:LineStringType" substitution-
  Group="gml:_Geometry" />

= <complexType name="LineStringType">
  = <annotation>
    <documentation>A LineString is defined by two or more co-
      ordinate tuples, with linear interpolation between
      them.</documentation>
  </annotation>
  = <complexContent>
    = <extension base="gml:AbstractGeometryType">
      = <sequence>
        = <choice>
          <element ref="gml:coord" minOccurs="2"
            maxOccurs="unbounded" />
          <element ref="gml:coordinates" />
        </choice>
      </sequence>
    </extension>
  </complexContent>
</complexType>

```

Anh. B.4: Geometry Schema für den LineString<sup>69</sup>

---

<sup>68</sup> Vgl. Geometry Schema.

<sup>69</sup> Vgl. Geometry Schema.

```
<element name="Polygon" type="gml:PolygonType" substitution-
  Group="gml:_Geometry" />

= <complexType name="PolygonType">
  = <annotation>
    <documentation>A Polygon is defined by an outer boundary
      and zero or more inner boundaries which are in turn de-
      fined by LinearRings.</documentation>
  </annotation>
  = <complexContent>
    = <extension base="gml:AbstractGeometryType">
      = <sequence>
        = <element name="outerBoundaryIs">
          = <complexType>
            = <sequence>
              <element ref="gml:LinearRing" />
            </sequence>
          </complexType>
        </element>
        = <element name="innerBoundaryIs" minOccurs="0"
          maxOccurs="unbounded">
          = <complexType>
            = <sequence>
              <element ref="gml:LinearRing" />
            </sequence>
          </complexType>
        </element>
      </sequence>
    </extension>
  </complexContent>
</complexType>
```

Anh. B.5: Geometry Schema für das Polygon<sup>70</sup>

---

<sup>70</sup> Vgl. Geometry Schema.

```
<element name="MultiPoint" type="gml:MultiPointType" substitution-  
  Group="gml:_Geometry" />  
  
= <complexType name="MultiPointType">  
  = <annotation>  
    <documentation>A MultiPoint is defined by one or more  
      Points, referenced through pointMember ele-  
      ments.</documentation>  
  </annotation>  
  = <complexContent>  
    = <restriction base="gml:GeometryCollectionType">  
      = <sequence>  
        = <element name="pointMember" maxOc-  
          curs="unbounded">  
          = <complexType>  
            = <sequence>  
              <element ref="gml:Point" />  
            </sequence>  
          </complexType>  
        </element>  
      </sequence>  
    </restriction>  
  </complexContent>  
</complexType>
```

Anh. B.6: Geometry Schema für den MultiPoint<sup>71</sup>

---

<sup>71</sup> Vgl. Geometry Schema.

## **Abschließende Erklärung**

Ich versichere hiermit, dass ich meine Diplomhausarbeit "Repräsentations- und Abfragemechanismen für geographische Informationen im Web" selbstständig und ohne fremde Hilfe angefertigt habe, und dass ich alle von anderen Autoren wörtlich übernommenen Stellen wie auch die sich an die Gedankengänge anderer Autoren eng anlehnenden Ausführungen meiner Arbeit besonders gekennzeichnet und die Quellen zitiert habe.

Münster, den 20. Dezember 2002