

Suchmaschinen mit räumlichem Bewusstsein

Monika Sester und Frauke Heinzle
Institut für Kartographie und Geoinformatik
Universität Hannover

Internetsuchmaschinen basieren darauf, dass ein eingegebener Suchbegriff mit einem Datensatz auf Internetseiten im verteilten World-Wide-Web übereinstimmt. Mit diesem reinen Keyword-Matching kann man allerdings nur die Information erhalten, die explizit gespeichert ist. Das von der EU geförderte Projekt SPIRIT hat zum Ziel, im Internet auch Anfragen zu stellen, die mit räumlichen Daten zusammenhängen: Dann wird es möglich sein, nach Hotels im Norden Hannovers zu suchen, und auch Antworten zu Hotels in Garbsen (einer Stadt nord-westlich von Hannover) zu erhalten. D.h. es sind sowohl Abfragen über räumliche Objekte, als auch über räumliche Zusammenhänge möglich.

1 Einführung und Übersicht

Die Nutzung von Suchmaschinen im Internet ist in vielen Bereichen unseres täglichen Lebens zur Selbstverständlichkeit geworden. Ein großer Bereich der Informationsrecherche bezieht sich u.a. auf touristische Anwendungen, wo es um Reiseplanung, Hotelbuchungen etc. geht. Solche Anwendungen haben bereits eine große Verbreitung erreicht. Bei dieser Fragestellung spielt der Raumbezug eine große Rolle, da typischerweise *Hotels in zentraler Lage, in der Nähe des Flughafens* etc. gesucht werden.

Heutige Suchmaschinen lösen diese Aufgabe, indem sie nach diesen Begriffen in Web-Seiten suchen. Ein reines Keyword-Matching überprüft, ob die gesuchten Begriffe in einem Internet-Dokument vorkommen und zeigt die Treffer dann an – wobei die wahrscheinlichsten Treffer vorne in der Liste erscheinen.

Die Suche nach Informationen über die Stadt *London* wird sehr viele Treffer liefern, da alle Texte auftauchen, in denen dieser Begriff vorkommt. Der semantische Kontext der Suche und der Dokumente selbst kann noch nicht spezifiziert werden – es entscheidet allein das Vorhandensein bzw. die Häufigkeit des Auftretens des Begriffs. Im Bereich des sog. Semantic Web [Berners-Lee et al. 2001] versucht man daher, die Semantik eines Begriffes mit zu beschreiben. Fernziel ist dabei, dass intelligente Agenten im Netz relativ selbstständig nach relevanter Information suchen können, da der jeweilige Kontext der Suchanfragen bzw. der Webseiten durch die semantischen Annotationen eindeutig gemacht wurden.

Ziel des EU-Projekts SPIRIT [Jones et al., 2002] ist es, eine Suchmaschine mit räumlichem Bewusstsein zu entwickeln, welche die Semantik räumlicher Begriffe „versteht“.

Eine Teilaufgabe des SPIRIT-Projektes besteht darin, raumbezogene Datensätze und ihren Inhalt im Internet „sichtbar“ zu machen. Der Grund hierfür ist zweierlei: einerseits existiert eine Fülle von räumlichen Datensätzen. Allerdings liegen diese Datenbestände typischerweise auf den Datenservern ihrer Hersteller und sind daher nicht zugänglich für einen Zugriff von außen. Zum zweiten ist der eigentliche Inhalt der Daten größtenteils graphisch beschrieben, d.h. in den Datensätzen ist die meiste Information nur implizit enthalten. Diese implizite Information kann allerdings für eine Fülle von Fragestellungen eingesetzt werden – auch für solche, für die ein Datensatz ursprünglich gar nicht erfasst wurde. Dies führt somit zu zwei Zielen:

- a. Sichtbarmachen der raumbezogenen Datenbestände im Internet bedeutet, dass Information über die Daten (Metadaten) im Internet bekannt gemacht werden. Diese Metadaten müssen selbstverständlich aussagekräftig genug sein, um hinreichende

Information über die Daten zu beinhalten. Typischerweise werden diese Metadaten aufwändig von Hand erhoben.

- b. Das ambitioniertere Ziel liegt darin, auch die Inhalte der Daten im Internet sichtbar zu machen, d.h. auch implizite Zusammenhänge in den Daten offen zu legen. Diese Zusammenhänge sollen mittels Data Mining Methoden automatisch aus den Geodaten abgeleitet werden. Hierunter fallen beispielsweise Informationen wie *gute Wohngegend*, *angenehmer Fahrradweg*, etc. d.h. Information, die in den Daten typischerweise nicht kodiert ist, sich einem intelligenten Kartennutzer allerdings direkt erschließt.

Im Beitrag werden die Ziele von SPIRIT kurz dargestellt und die Komponenten eines solchen Systems vorgestellt. Es erfolgt ein Fokus auf die Bereitstellung der Ontologie, d.h. des notwendigen Hintergrundwissens über räumliche Objekte und ihre Zusammenhänge.

2 Räumliche Ontologie

Das Basiswissen um räumliche Objekte und Zusammenhänge muss dem System vorliegen, es wird in Form von Ontologien bereitgestellt ([Jones et al., 2001], [Bernard et al., 2003]). Es existieren bereits Repräsentationen geographischer Ontologien, welche im Kontext von sogenannten Thesauri oder Gazetteers entwickelt wurden. Hierunter zählen z.B. der Getty Information Institute's Thesaurus of Geographic Names [TGN 2002] oder die Alexandria Digital Library gazetteer [ADL, 2002]. Im TGN werden Orte hierarchisch gegliedert, d.h. über Länder, administrative Einheiten, etc. Der Raumbezug wird entweder über Punkte oder Polygone (footprint) gegeben. Abbildung 1 zeigt eine Darstellung der Kodierung der Stadt Florenz im TGN.

<p>Record ID: [7000457]</p> <p>Coordinates: Lat: 43 47 N Long: 011 15 E</p> <p>Note: Was a Roman military center at the head of navigation on the Arno river</p> <p>Hierarchical Position: ...Europe.....(continent) Italia.....(nation) Toscana.....(region) Firenze.....(province)</p> <p>Names: Firenze (C,V, preferred) Florence (C,O, prefEnglish) Firencia (C,O)</p> <p>Place Types: inhabited place (C), city (C), regional capital (C), provincial capital (C), commune (administrative) (C), river settlement (C), tourist center (C)</p> <p>Sources and Contributors: Firenza.... <i>Companion Guide: Florence (1979), 14 [VP]</i> Firenze.... <i>Columbia Lippincott Gazetteer (1961) [BHA]; Webster's Geographical Dictionary (1984) [GRLPA]; Times Atlas of the World (1992), 66 [VP]; Companion Guide: Florence (1979), 62 ff. [FDA]</i></p>

Abbildung 1: Beispiel: Beschreibung der Stadt Florenz im TGN.

SPIRIT zielt auf die Bereitstellung einer Ontologie ab, welche neben den räumlichen Begriffen auch Beziehungen mit einbezieht. Es ergeben sich als Aufgaben von SPIRIT [Fu et al. 2003]:

- Identifizieren eines Ortes bzw. einer räumlichen Beziehung in einem Web-Dokument bzw. einer Anfrage.
- Auffinden von Web-Dokumenten, welche Alternativen zum angegebenen Namen beinhalten.
- Auffinden von Web-Dokumenten, welche sich auf Orte innerhalb oder in der Nähe des angegebenen Namens befinden.
- Unterscheidung unterschiedlicher Ortstypen.
- Schnelles Auffinden relevanter Web-Dokumente mittels effektiven Indexmechanismen.
- Sortieren der gefundenen Dokumente anhand der Qualität der Übereinstimmung mit den angegebenen räumlichen und thematischen Komponenten der Anfrage.

Das Schema der Ontologie wurde in Fu et al. [2003] definiert. In Abbildung 2 ist das UML Klassendiagramm angegeben.

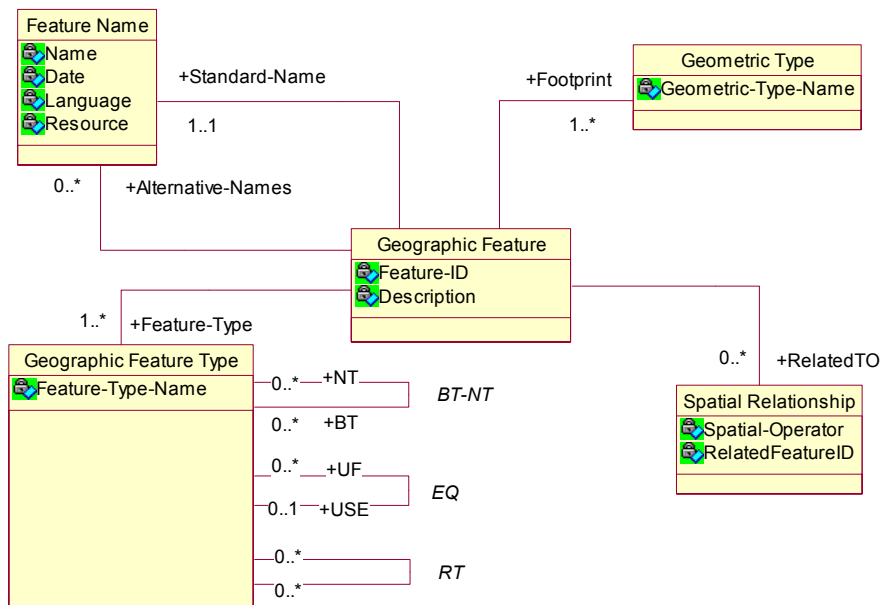


Abbildung 2 Schema der SPIRIT Ontologie

Für jedes Geoobjekt wird angegeben: eine eindeutige Objekt-Identifikation, ein Name, das Erfassungsdatum, sowie eine zugehörige Sprache; alternative Namen geben Varianten an, unter denen das Objekt auch auftreten kann. Schließlich wird jedes Objekt einer Objektart zugeordnet (z.B. Fluss, Stadt). Ein wichtiges Element ist weiterhin der räumliche Bezug in Form eines *footprints*, welcher als Punkt, als umschließendes Rechteck oder als genauere Geometrie gegeben sein kann. Von besonderer Bedeutung sind die räumlichen Beziehungen (*SpatialRelationship*), welche explizit modelliert werden. Es sind die topologischen Beziehungen: innerhalb, enthält, überlappend, angrenzend.

Die Ontologie stellt einerseits das formale Schema dar; zum anderen muss sie aber auch konkrete Instanzen beinhalten. Dieses Erstellen der SPIRIT-Ontologie wird mit Hilfe vorhandener Geodatenätze durchgeführt. Derzeit wird der europäische SABE-Datenatz integriert, der administrative Grenzen in Europa enthält

[http://www.eurogeographics.org/eng/04_sabe.asp]; ferner sollen weitere Daten aus den Partnerländern exemplarisch eingebunden werden (speziell topographische Daten). Schließlich werden weitere Informationen für Touristenapplikationen integriert.

Es stellt sich allerdings die Frage, ob alle Information, die für die möglichen räumlichen Anfragen relevant sind, sich direkt in den existierenden Datenquellen ermitteln lassen? - Es kann beobachtet werden, dass selbst in digitalen raumbezogenen Datenbeständen ein Großteil der Information lediglich implizit in den Daten vorhanden ist. Diese Information kann über Interpretationsverfahren erschlossen und somit der Ontologie zugeführt werden. Somit werden in SPIRIT raumbezogene Daten in zweierlei Weise direkt eingesetzt:

1. Direkte Suche nach raumbezogenen Daten (z.B. Suche nach *Landnutzungskartierung in Südengland*).
2. Nutzung der Information in den raumbezogenen Datenbeständen zur Ausstattung der Ontologie
 - a. Explizite Information: diese kann direkt in die Ontologie übertragen werden.
 - b. Implizite Information: diese muss zunächst aus den Daten extrahiert werden.

Am ikg stehen die Fragen 1 und 2b im Vordergrund. Im folgenden werden zunächst die Metadaten beschrieben, die für Frage 1 nötig sind, sowie Verfahren zu deren Extraktion; anschließend werden Basisinformationen und Extraktionsansätze für Frage 2b präsentiert.

Eine sinnvolle Bearbeitung von sehr vielen räumlichen Datensätzen kann nur dann erfolgen, wenn die Prozesse der Informationsextraktion weitestgehend automatisch ablaufen, daher wird ein hoher Grad an Automation angestrebt.

3 Automatische Annotierung raumbezogener Datensätze

In Geodaten (z.B. ATKIS oder Biotopkartierung) ist eine Fülle an Information direkt und explizit in Form der einzelnen Datenobjekte gespeichert. Die Information darüber, welchen Inhalt ein spezieller Datensatz hat, wird typischerweise in Metadaten abgelegt. Es existiert eine Reihe von Portalen, welche für die Suche nach raumbezogenen Daten ausgelegt sind. In Deutschland sind dies z.B. GeoMis.Bund (www.geomis.bund.de), TerraCatalog (www.conterra.de), InGeoIC (<http://www.ingeoic.de/>), GeoMDK (<http://www.geomdk.niedersachsen.de/>). Dort sind Datensätze über Metadaten beschrieben und somit – anhand der vorgegebenen Schlüsselbegriffe auch auffindbar.

Das Problem stellt dabei allerdings die meist manuelle Zuordnung der Metadaten-Elemente (Tags) dar. Es existieren Tools, welche eine manuelle Erfassung unterstützen. Der Aufwand zur Erhebung dieser Metadaten ist jedoch dennoch sehr hoch, daher wären Verfahren ideal, die diesen Prozess automatisieren. Im folgenden Abschnitt werden zunächst Metadaten-Elemente vorgestellt, welche für ein Auffinden im Internet elementar wichtig sind. Diese wurden in Anlehnung an internationale Standards zusammengestellt. Ziel hierbei ist es jedoch, die wichtigsten Elemente festzulegen, sowie Methoden für ihre automatische Extraktion anzudeuten.

3.1 Metadaten

Die in SPIRIT verwendeten Metadaten lehnen sich sehr stark an existierende internationale Standards an [ISO/TC211 2003]. Die folgenden Metatags sind für SPIRIT von essentieller Bedeutung:

1. **Räumliche Ausdehnung / Spatial extent:** es gibt unterschiedliche Arten, die räumliche Ausdehnung von Objekten zu definieren, wovon die wesentlichen folgende sind:
 - a. Ein Minimal einschließendes Rechteck (bounding box)

- b. Ein geographischer Name
 - c. Die Objektgeometrie
2. **Stichworte / Keywords:** dieses Metadatenelement beschreibt den Inhalt der Daten sowie den Anwendungsbereich. Die Stichworte können auf unterschiedliche Art festgelegt werden:
 - a. Eine offene Liste mit beliebigen Stichworten
 - b. Eine offene Liste mit beliebigen Stichworten, welche durch natürlich-sprachliche Erklärungen ergänzt ist
 - c. Eine Auswahl von Stichworten, welche beispielsweise aus der ISO-Liste MD_TopicCategoryCode stammt oder aus der SPIRIT Ontologie.
 3. **Kontakt:** In jedem Datensatz sollte spezifiziert sein, über welchen Kontakt der Datensatz selbst beziehbar ist (z.B. über eine URL oder eine Adresse. Dort könnten Modalitäten beschrieben sein, wie der Datensatz beschafft werden kann (Person bzw. Download-Adresse).
 4. **Scale/Resolution (optional):** eine Auflösung (bzw. ein Maßstab) gibt an, auf welche Auflösungsebene sich die Daten des Datensatzes beziehen.

3.2 Automatische Extraktion von Metadaten aus ArcView Shapefiles

Das Shape-Format von ESRI ist allgemein bekannt und les- bzw. auswertbar. Folgende Informationen sind u.a. daraus zu entnehmen:

- bounding box (Ausdehnung entsprechend kleinster und größter Koordinate)
- Anzahl der geometrischen Elemente
- Typ der geometrischen Elemente (Punkt, Linie, Polygon, 2D, 3D)
- Attributinformation (aus den dazugehörigen dbf-Files), Struktur der Attribute (Name, Typ (String, Integer, Double etc.), Spaltenweite, Dezimalstellen)

Diese Informationen sind wichtig für die Interpretation des geometrischen Aspekts eines Datensatzes. Allerdings sagen diese Daten nur wenig über Inhalte aus, vor allem dann, wenn die Attributnamen auch noch durch Nummern oder sonstige Codes verschlüsselt sind, wie es z.B. in ATKIS üblich ist.

Um einem Internetnutzer die „richtige“ Information liefern zu können, müsste man zumindest wissen, welche Daten in dem Datensatz verschlüsselt sind. Also nicht nur das Wissen, dass darin Linien enthalten sind, sondern vielmehr, ob diese Linien ein Straßennetz, Telefonleitungen, Begrenzungen eines Gebietes oder Höhenlinien darstellen.

Hierfür werden die Shapefiles und die dazugehörigen ArcView-Projekte untersucht, um automatisch entsprechende Keywords abzuleiten. Folgendes Beispiel soll dies belegen:

ATKIS-Datensatz – Name des shape-Files: F104_LT.shp

Auszug aus dem dazugehörigen dbf-File

AOBJID	OBJ	TEIL	OBJART	AKT	OART_ATYP
B01HW10	B01HW10	001	3102		3102
B01HXUA	B01HXUA	001	3101		3101_1307
B01HXUG	B01HXUG	001	3102		3102

Aus obigen Daten ist nicht ersichtlich, dass es sich hier um ein Straßennetz handelt. Dieses ist in Abbildung 3 (links) abgebildet. Rechts sind die zugehörigen Metadaten dargestellt, welche automatisch abgeleitet wurden. Abbildung 4 zeigt ein Beispiel für Gewässer.

Ein weiteres Beispiel wurde aus einem französischen Datensatz generiert. Es handelt sich dabei um den topographischen Datenbestand BDTopo (siehe Abbildung 5).



```

<Metadata>
  <Name> Straßenverkehr (104) - Objektteil-Linien </Name>
  <Path> M:\heinzle\lckern-Waltrup\arcview\F104_LT.shp </Path>
  <CreateDate> Mittwoch, 30. Oktober 2002 14:08:47 </CreateDate>
  <Keywords>
    25      Straße
    29      Bundesautobahn
    52      Bundesstraße
    163     Landesstraße, Staatsstraße
    80      Forststraße
    865     Gemeindestraße
    357     Weg
    12      Fahrbahn
  </Keywords>
  <Number of Entities> 1583 </Number of Entities>
  <ShapeType> Polylines </ShapeType>
  <Min X> 2585164.000 </Min X>
  <Min Y> 5716250.000 </Min Y>
  <Max X> 2594000.000 </Max X>
  <Max Y> 5724000.000 </Max Y>
</Metadata>

```

Abbildung 3: Straßendatensatz (links) und zugehörige Metadaten.



```

<Metadata>
  <Name> Wasserflächen (111) - Objektteil-Linien </Name>
  <Path> M:\heinzle\lckern-Waltrup\arcview\F111_LT.shp </Path>
  <CreateDate> Mittwoch, 30. Oktober 2002 14:08:47 </CreateDate>
  <Keywords>
    221     Strom, Fluß, Bach
    56      Graben, Kanal (Wasserwirtschaft)
  </Keywords>
  <Number of Entities> 277 </Number of Entities>
  <ShapeType> Polylines </ShapeType>
  <Min X> 2587421.800 </Min X>
  <Min Y> 5716972.200 </Min Y>
  <Max X> 2594000.000 </Max X>
  <Max Y> 5723128.100 </Max Y>
</Metadata>

```

Abbildung 4: Gewässerelemente eines Datensatzes; rechts: zugehörige Metadaten.

```

<Metadata>
  <Name> Etablissement.shp </Name>
  <Path> d:\donnees\bd carto@version objet/bdc09_av/toponymie/etablissement.shp </Path>
  <CreateDate> Mittwoch, 19. November 2003 17:54:45 </CreateDate>
  <Keywords>
    0      Aérogare
    1      Enseignement
    1      Hôpital
    0      Hôtel départemental
    0      Hôtel régional
    0      Préfecture
  </Keywords>
  <Number of Entities> 2 </Number of Entities>
  <ShapeType> Points </ShapeType>
  <Min X> 540622.000 </Min X>
  <Min Y> 1790201.000 </Min Y>
  <Max X> 541830.000 </Max X>
  <Max Y> 1790733.000 </Max Y>
</Metadata>

```

Abbildung 5: automatisch extrahierte Metadaten aus einem Datensatz des französischen BDTopo.

Zur Ermittlung der Keywords werden alle zugänglichen Dateien untersucht. So werden Textfiles überprüft und Straßennamen oder Regionsbezeichnungen gefunden. Die Legenden, welche im Projektfile abgespeichert sind, geben häufig Auskunft über die abgespeicherten Elemente, oder auch die dazugehörigen dbf-Files mit den Attributnamen.

Angaben zum Maßstab können entweder direkt einer Beschreibung des Datensatzes entnommen werden (sofern diese vorhanden ist). Sie können jedoch auch aus der Geometrie der Objekte genähert abgeleitet werden: die Punktdichte einer Linie gibt beispielsweise Aufschluss darüber, für welchen Maßstabsbereich ein Geobjekt erfasst worden ist.

Mithilfe der Keywords können dem Datensatz bereits wichtige Informationen entnommen werden. Allerdings sind dadurch noch keine Kenntnisse über die Verteilung bzw. Lage der Geometrieelemente (sozusagen ein Bild der Dinge), deren Beziehungen untereinander, deren Häufung an bestimmten Orten u.ä. vorhanden. Dies führt uns zum nächsten Schritt, um Geodatensätze zu erschließen und im Internet, speziell für Suchmaschinen, zugänglich zu machen: der Extraktion impliziter Information.

4 Implizite Daten in Geodaten

4.1 Extraktion impliziter Information aus Geodatensätzen

Das Aufdecken impliziter Zusammenhänge in den Daten ist ein Prozess, den ein geübter Kartennutzer ohne Probleme durchführt: er erkennt einen schönen Wanderweg, da er an Flüssen entlang und durch Wälder führt, eine gute Wohngegend oder auch ein Industriegebiet, welches sich für eine Niederlassung einer Filiale eignen würde. D.h. diese Information ist in den Daten vorhanden, jedoch typischerweise nicht explizit. Das Erschließen dieser Information bedarf Interpretationsmethoden, ähnlich wie sie im Bereich der automatischen Bildanalyse eingesetzt werden (z.B. [Haralick and Shapiro 1999]). Weiterhin können Techniken der algorithmischen Geometrie (z.B. [de Berg et al. 2000]) bzw. des räumlichen Data Minings genutzt werden (z.B. [Kopierski and Han 1995 9]). Im folgenden Abschnitt werden einige für SPIRIT relevante räumliche Konzepte vorgestellt, parallel dazu die Eigenschaften, aus denen sie sich erschließen lassen.

4.1.1 Klassifikation in bedeutende Städte

Die Bedeutung einer Stadt ist üblicherweise nicht direkt im Datenbestand abgelegt (außer implizit über ihre Größe bzw. Klassifikation in Kreis- bzw. Hauptstadt). Die relative Bedeutung kann aber an verschiedenen Faktoren abgelesen werden, welche im folgenden exemplarisch aufgezählt werden.

Gegebene Information in den Daten	Nötige Eigenschaften
Siedlungsgebiete	Ausdehnung, Größe, Dichte, Kategorien (Wohnbaufläche, Industriegebiet, etc.)
Straßen	Straßenklassen; Kreuzungen
Eisenbahn, Bahnhof	Kreuzungen, Bahnhöfe, räumliche Lage zur Stadt (Rand, Mitte, ...)
Öffentlicher Nahverkehr	Art, Anzahl und Dichte der Stationen, ...
Öffentliche Gebäude	Art, Anzahl, Einzigartigkeit in gewisser Umgebung, ...
Schulen, Universitäten, sonstige	Anzahl, Einzigartigkeit, Einzugsgebiet, ...

Bildungseinrichtungen	
Kirchen	Anzahl, Konfession, Verteilung
Einkaufszentren	Größe, Einzugsgebiet, Produkte
Marktplatz	Anzahl
Verwaltungsgrenzen	Ausdehnung, Größe, Arten
Hafen	Existenz, räumliche Lage zur Stadt
Weitere Städte	Vergleich, Ranking von Städten gleicher Kategorie in ihrer Nachbarschaft (ermittelt z.B. durch Voronoi-Diagramm)

4.1.2 Bestimmung des Stadtzentrums

Die Information über die genaue Lage des Stadtzentrums ist üblicherweise nicht im Datensatz angegeben. Sie lässt sich jedoch aus folgenden Informationen ermitteln.

Gebäude	Typ, Anzahl, Dichte, Unterschiedlichkeit in räumlicher Nähe
Siedlungsgebiet	Ausdehnung, Dichte
Straßen	Kreuzungen
Öffentliche Gebäude, Handel	Vorhandensein, relative Dichte
Marktplatz	Räumliche Lage
Haltestellen des öffentlichen Verkehrs	Kreuzungen, Umsteigemöglichkeiten, ...
Industriegebiet	Nicht vorhanden im Stadtzentrum

4.1.3 Attraktive touristische Gebiete

Für Tourismus attraktive Gebiete zeichnen sich durch die im folgenden aufgezählten Informationsquellen aus.

Hotel, Pension, Jugendherberge	Anzahl, Verteilung
Kulturelle Einrichtungen	Typ, Verteilung, Dichte
Waldgebiete	In der Stadt, außerhalb, Dichte,
Restaurants	Anzahl, Verteilung
Naherholungsgebiete	Anzahl, Verteilung, Größe
Sight-Seeing	Wichtige Objekte

4.2 Automatische Verarbeitung

Für eine automatische Klassifikation dieser Kategorien müssen nun folgende Fragen geklärt werden:

1. Ermittlung der relevanten Eigenschaften (Operationen)
2. Kombination dieser Eigenschaften

Was die Ermittlung der Eigenschaften betrifft, so lassen sich einige durch einfache GIS-Operationen durchführen – etwa das Auszählen des Vorkommens eines bestimmten Objektes. Schwieriger sind Operationen, die sich rein auf die Geometrie beziehen, speziell die Ermittlung der Verteilung bzw. Dichte. Diese kann z.B. durch lokale Analysen von Abständen durchgeführt werden – allerdings sind hier typischerweise Schwellwerte erforderlich, die schwierig zu bestimmen sind. Im Bereich des Clusterings wurden Verfahren entwickelt, welche auf

Nachbarschaftsgraphen basieren; Anders [2003] stellt eine Methode vor, welche keine Schwellwerte benötigt.

Die Kombination der Eigenschaften stellt ein weiteres großes Problem dar: sie können zwar über logische Kombinationen zusammengeführt werden. Allerdings sind die oben gelisteten Eigenschaften u.U. nicht gleichwertig; weiterhin sind manche sicherlich zwar wichtig, aber nicht essentiell nötig, um ein bestimmtes Konzept identifizieren zu können. Daher sind hier Verfahren einzusetzen, welche die Bedeutung und die Genauigkeit der Eingangsgrößen mit berücksichtigen (etwa über Wahrscheinlichkeiten oder Fuzzy-Logik).

5 Zusammenfassung und Ausblick

Im Beitrag wurde eine wichtige Grundlage vorgestellt, welche für die Entwicklung von „semantischen“ räumlichen Suchmaschinen nötig ist: die Bereitstellung einer räumlichen Ontologie, in der die Objekte und ihre Beziehungen beschrieben sind, und somit von Suchmaschinen genutzt werden können. Diese Konzepte können aus räumlichen Datenbeständen ermittelt werden. Dafür wurde im ersten Teil ein Vorschlag für eine Annotierung von räumlichen Datensätzen mittels einer Auswahl wichtiger Metadaten gemacht, welche sich mit einfachen Methoden automatisch ermitteln lassen, wodurch ein aufwändiger manueller Prozess entfällt. Im zweiten Teil schließlich wurden interessante räumliche Konzepte vorgestellt, welche implizit in Geodaten vorhanden sind, aber mittels räumlicher Interpretationsverfahren erst erschlossen werden müssen, um sie für Suchmaschinen nutzbar zu machen. Zukünftige Arbeiten konzentrieren sich speziell auf diesen Punkt, indem allgemeine Methoden zur Ermittlung räumlicher Zusammenhänge entwickelt werden, sowie Modelle erarbeitet werden, die die Kombination elementarer Eigenschaften der Konzepte ermöglichen.

6 Literatur

ADL [2002], Alexandria Digital Library Project, <http://www.alexandria.ucsb.edu/>, 2002.

Anders, K.-H. [2003]: Parameterfreies hierarchisches Graph-Clustering Verfahren zur Interpretation raumbezogener Daten, Dissertation Universität Stuttgart, 2003.

Arampatzis, A. & van Kreveld, M. [2003]: Simple and Useful Similarity Measures. D9:5102. SPIRIT Project. September 2003.

de Berg, M. M. van Kreveld, M. Overmars & O. Schwarzkopf (2000), *Computational Geometry -- Algorithms and Applications*, Springer, Heidelberg

Bernard, L., U. Einspanier, S. Haubrock, S. Hübner, W. Kuhn, R. Lessing, M. Lutz, U. Visser (2003): [Ontologies for Intelligent Search and Semantic Translation in Spatial Data Infrastructures](#). Photogrammetrie-Fernerkundung-Geoinformation 6/2003: 451-461.

Berners-Lee, Tim, James Hendler & Ora Lassila [2001], The Semantic Web, Scientific American, Mai 2001.

Fu, G., A. Abdelmoty & Ch. Jones [2003]: Design of a Geographical Ontology, Report D5 3101, SPIRIT project, <http://www.geo-spirit.org/>.

Haralick, R.M. & L.G. Shapiro [1993], *Computer and Robot Vision*, Addison-Wesley, Vol. I & II.

ISO/TC-211 (2003): *Text for FDIS 19115 Geographic information - Metadata. Final Draft Version*, International Organization for Standardization.

Jones, C.B., H. Alani and D. Tudhope [2001], Geographical information retrieval with ontologies of place. D. Montello (ed), *Spatial Information Theory Foundations of Geographic Information Science*, COSIT 2001, Lecture Notes in Computer Science 2205, 2001, 323-335.

Jones, C.B. et al. [2002] : Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project". *Proc. of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.387 – 388, Tampere, Finland, August 11-15, 2002.

Koperski, K. & Han, J. (1995), Discovery of Spatial Association Rules in Geographic Databases, *in: M. J. Egenhofer & J. R. Herring, eds, 'Advances in Spatial Databases '95', Vol. 951 of Lecture Notes in Computer Science*, Springer Verlag, Heidelberg, pp. 47–66.

TGN [2002], Getty Thesaurus of Geographic Name,
<http://www.getty.edu/research/tools/vocabulary/tgn/>, 2002.

Dank: Die Arbeit entstand im Rahmen des EU-Projektes SPIRIT. Die ATKIS Daten wurden uns von der Landesvermessung und Geobasisinformation Niedersachsen (LGN) zur Verfügung gestellt.