

## Georeferenzierung von Informationen auf dem Internet

Wolf-Fritz Riekert, FAW Ulm<sup>1</sup>

Die Zunahme von Informationen, die auf dem Internet angeboten werden, hält unvermindert an. Dieser Prozeß ist im Hinblick auf die Bildung einer „Informationsgesellschaft“ als sehr positiv zu sehen; zugleich fällt es den Informationsnutzern aber zunehmend schwer, sich in dem schier unüberschaubaren Angebot von Informationsdiensten zu orientieren. Diese Orientierungsproblematik, auch als „Lost in Hyperspace“-Syndrom beklagt, macht besondere Systeme und Werkzeuge erforderlich, die eine Hilfe bei der Suche nach Informationen anbieten.

So ermöglichen die verschiedenen im Internet angebotenen „Suchmaschinen“ die Suche nach Hypertext-Seiten im World-wide Web (WWW) in Form einer Volltextrecherche. Diesen Suchmaschinen liegt ein Volltextindex zugrunde, der potentiell das gesamte WWW umfaßt und der durch vollautomatische Programme, sog. „Robots“, gepflegt wird, die selbständig das Geflecht von Verweisen im WWW verfolgen und eine Indizierung der gefundenen Hypertextseiten vornehmen. Da über Hypertext-Seiten vielfältige Informationsangebote erreichbar sind, die außer multimedialen Dokumenten auch Daten (inklusive Geodaten) und Dienstprogramme, z.B. zur Datenbankabfrage umfassen, ermöglichen Suchmaschinen über eine einfache Stichwortsuche den Zugang zu einem breiten Spektrum von Informationen.

Ein Nachteil der Volltextrecherche ist allerdings die ausschließlich textuelle Auswertung der Stichwörter. Eine inhaltliche Interpretation der Stichwörter ist der Suchmaschine nicht möglich. So würde beispielsweise über die Angabe der Stichwörter „Südfrucht“ und „Bodensee“ ein im WWW eventuell vorhandener Artikel über den Anbau von Apfelsinen auf der Insel Mainau nicht gefunden werden, da die Suchmaschine über keinerlei Wissen darüber verfügt, daß Apfelsinen Südfrüchte sind und die Insel Mainau im Bodensee liegt.

Die Verschlagwortung von Informationen über einen Volltextindex reicht also nicht für alle Anwendungsgebiete aus. Es werden deshalb in semantischer Hinsicht tiefer greifende Kriterien für die Beschreibung von Informationen benötigt. Für diesen Zweck wurden Metainformationssysteme entwickelt, die eine Katalogisierung und Recherche von Informationen nach derartigen Kriterien erlauben. Bei der Konzeption von derartigen Metainformationssystemen hat sich gezeigt, daß es sinnvoll ist, insbesondere drei Arten von Beschreibungselementen für Informationen zu unterscheiden, nämlich *Zeitbezüge*, *Fachbezüge* und *Raumbezüge*. Die im obigen Beispiel angeführte Recherche zielt auf Informationen mit Fachbezug „Südfrucht“ und Raumbezug „Bodensee“ ab. Ein Zeitbezug ist in dieser Anfrage nicht spezifiziert; ein solcher könnte beispielsweise durch die zusätzliche Angabe einer Jahreszahl hergestellt werden. Ein Beispiel für ein Metainformationssystem, das die Katalogisierung und Recherche von Informationsquellen mit den vorgenannten Beschreibungselementen erlaubt, ist der *Locator* des Umweltinformationsnetzes Deutschland GEIN (German Environmental Information Network); Abbildung 1 zeigt die Rechercheoberfläche des GEIN-Prototyps.<sup>2</sup>

---

<sup>1</sup> Dr. Wolf-Fritz Riekert, FAW Ulm, Postfach 2060, D-89010 Ulm, Email: riekert@faw.uni-ulm.de, Tel.: 0731 501 500, Fax: 0731 501 999.

<sup>2</sup> Der GEIN-Prototyp wurde vom FAW Ulm im Auftrag des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit und des Umweltbundesamts Berlin erstellt. Wir danken diesen Institutionen sowie dem Bund-Länder-Arbeitskreis Umweltinformationssysteme und dem Ministerium für Umwelt und Verkehr Baden-Württemberg für die Unterstützung dieser Arbeit.



**Abbildung 1:** Rechercheoberfläche des Prototyps für das Umweltinformationsnetz Deutschland GEIN. Unter den Überschriften Schlagwort, Region und Zeit können Informationsangebote nach Fachbezügen, Raumbezügen und Zeitbezügen recherchiert werden. In dieser Anfrage wird nach Informationen über Emissionen (*Fachbezug*) aus Baden-Württemberg (*Raumbezug*) im Jahr 1996 (*Zeitbezug*) gesucht. Das Ergebnis der Recherche wird in Form einer maussensitiven Liste von Informationsangeboten auf der rechten Seite des Recherchefensters angezeigt.

In Metainformationssystemen stellt sich die Formulierung und Auswertung von Zeitbezügen, also Zeitpunkten oder Zeitintervallen, auf die sich die fragliche Information bezieht, relativ einfach dar, da sie lediglich einfache syntaktische und numerische Verarbeitungen erfordert.

Schwieriger gestaltet sich die Behandlung von Fach- und Raumbezügen. Die einfachste und wirkungsvollste Möglichkeit, Fachbezüge festzulegen, geschieht mit Hilfe von Schlagwörtern. Wünschenswert ist allerdings ein kontrolliertes Vokabular für die Verwendung von Schlagwörtern. Dabei kann insbesondere auf Erfahrungen aus dem Bibliotheksbereich mit sogenannten *Thesauri* zurückgegriffen werden. Ein solcher Thesaurus besitzt zweierlei Funktionen. Zum einen stellt er einen Katalog dar, der alle verwendbaren Schlagwörter enthält, die für die Verschlagwortung und die Recherche von Informationen genutzt werden können. Zum anderen behandelt der Thesaurus diese Schlagwörter als *Begriffe*, die über *Beziehungen* zu einem semantischen Netz verknüpft sind. Im wesentlichen handelt es sich dabei um drei Beziehungstypen, nämlich die *Synonymbeziehung*, die *Ober-/Unterbegriff-Beziehung* und die *Beziehung zwischen verwandten Begriffen*. So läßt sich in einem solchen Thesaurus festlegen, daß „Südfrucht“ ein Oberbegriff des Begriffs „Orange“ und dieser wiederum ein Synonym des Begriffs „Apfelsine“ ist. Es bedarf keiner weiteren Erläuterung, daß der erste Teil der o.a. Recherche (Suche nach Informationen mit Fachbezug Südfrüchte) mit Hilfe des in einem Thesaurus repräsentierten begrifflichen Wissens durch ein intelligentes System automatisch lösbar ist.

Die Interpretation des zweiten Teils der o.a. Recherche (nach Informationen mit Raumbezug Bodensee) setzt eine sogenannte Georeferenzierung der Informationen voraus. Hierzu ist es erforderlich, daß jeder Information, die sich auf ein bestimmtes geographisches Gebiet bezieht, ein Raumbezug zugeordnet wird. Im allgemeinen Sinn sind Raumbezüge Beschreibungselemente für die Katalogisierung und Recherche von Informationen nach geographischen Kriterien. Raumbezüge lassen sich durch zwei wesentliche Merkmale charakterisieren, nämlich durch einen geographischen *Namen*, der zur textuellen Bezeichnung des Raumbezugs dient (z.B. im oben genannten Beispiel der Name „Mainau“), sowie durch eine sogenannte *Geometrie*, d.h. durch eine Menge geographischer Koordinaten, durch die Lage und Ausdehnung des Raumbezugs festgelegt ist. Die Festlegung des Raumbezugs einer Information bei deren Katalogisierung kann in einem Metasystem also prinzipiell auf zweierlei Weisen erfolgen, nämlich durch Angabe eines geographischen Namens oder durch die Festlegung von Koordinaten. Letzteres kann durch Eingabe numerischer Werte oder durch Zeigeaktionen mit Hilfe einer Maus erfolgen (Abbildung 2). Prinzipiell sind auch alle diese Möglichkeiten der Angabe von Raumbezügen bei der Recherche nach Informationen möglich.

Zur Verarbeitung von Raumbezügen ist geographisches Wissen erforderlich. Dieses läßt sich repräsentieren mit Hilfe eines sog. *Gazetteers*. Gazetteer heißt auf englisch Ortsverzeichnis; hier soll darunter eine Verwaltungsstruktur für Raumbezüge verstanden werden, ähnlich wie ein Thesaurus als Verwaltungsstruktur für Fachbezüge aufgefaßt werden kann. Raumbezüge lassen sich nach unterschiedlichen Typen unterscheiden; wichtige Beispiele sind Verwaltungseinheiten (d.h. Bundesländer, Landkreise, Gemeinden etc.), Postleitgebiete oder Kartenblätter. Raumbezüge besitzen einen eindeutigen Identifikator, der nicht notwendigerweise mit dem Namen des Raumbezugs identisch ist (z.B. Gemeindegrenznummer für Gemeinden). Wie bereits oben angeführt, ist ein Raumbezug darüber hinaus gegeben durch einen Namen und eine, in der Regel zweidimensionale, Geometrie.

Wie die Begriffe in einem Thesaurus sind auch Raumbezüge untereinander durch Beziehungen verknüpft. So ist es häufig möglich und sinnvoll, Raumbezüge hierarchisch zu ordnen, so z.B. Verwaltungseinheiten entsprechend der Verwaltungshierarchie oder Postleitgebiete entsprechend ihrer dezimalen Gliederung.



**Abbildung 2:** Selektion von Raumbezügen mit Hilfe der Maus im GEIN Prototyp. Auf dieser Abbildung ist die englischsprachige Oberfläche dargestellt.

Von besonderer Bedeutung sind geometrisch/topologische Beziehungen zwischen Raumbezügen. Diese müssen jedoch nicht explizit gespeichert werden, sondern lassen sich aus der Geometrie der Raumbezüge ableiten. Die wesentlichen geometrisch/topologischen Relationen sind dabei die Überlappung, das Ineinanderenthaltensein und die gegenseitige Berührung von Raumbezügen.

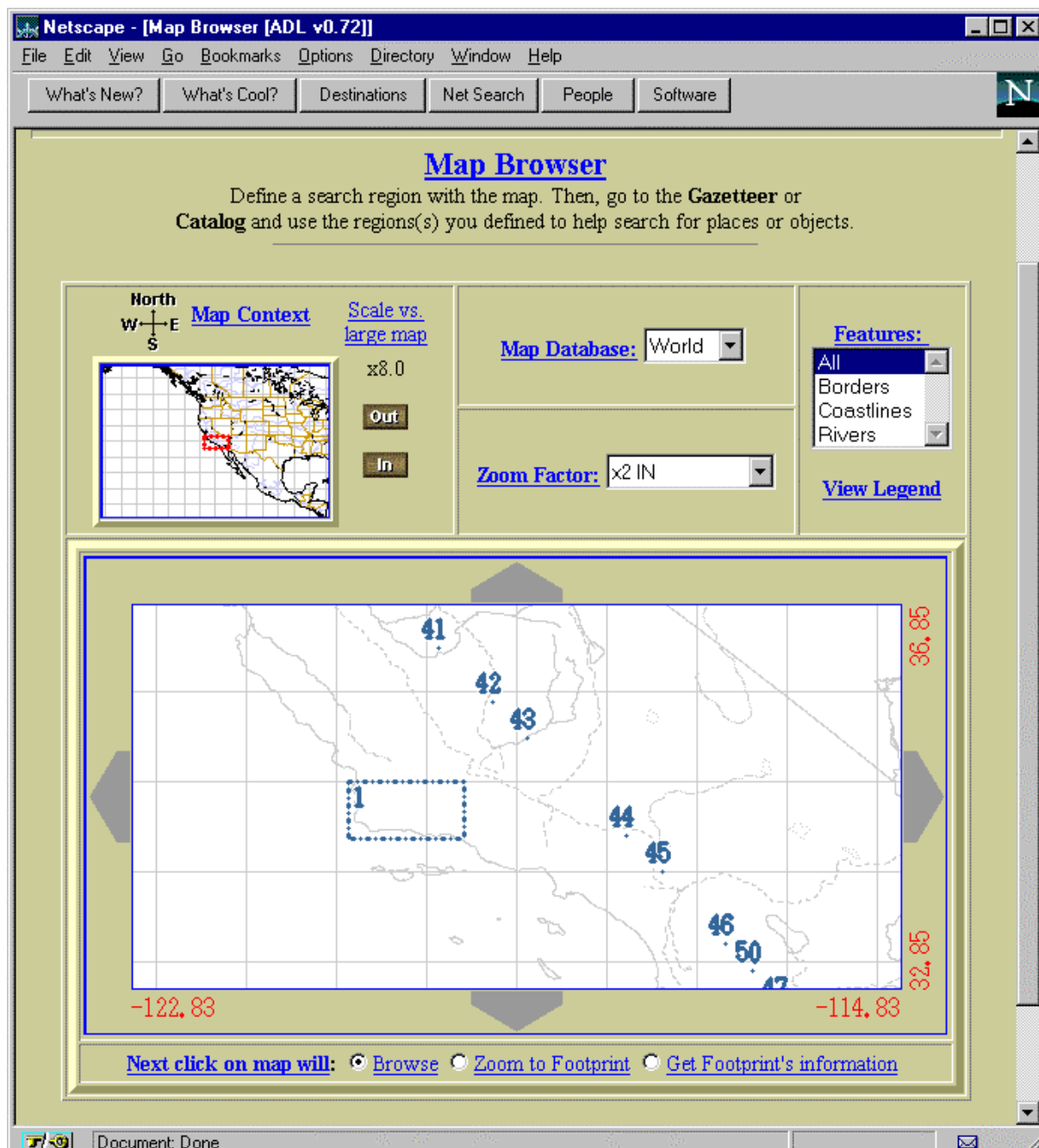
Die Geometrie eines Raumbezugs kann auf unterschiedliche Weise realisiert werden. Die einfachste Möglichkeit ist die Repräsentation der Geometrie in Form eines *umschließenden Rechtecks*. Dies bewirkt eine starke Vergrößerung der Geometrie der Raumbezüge, hat jedoch den Vorteil, daß für die Repräsentation des Gazetteers kein echtes GIS benötigt wird. Dies gilt auch für die etwas detailfreudigere Modellierung der Geometrien von Raumbezügen mit Hilfe von *Rasterzellen*. Hier wird das Plangebiet in eine hinreichend feine Rasterung von Planquadraten zerlegt. Diese Planquadrate lassen sich durchnummerieren und daher durch eine einfache Zahl identifizieren. Die Geometrie eines Raumbezugs ist gegeben durch die Menge aller

Planquadrate, die durch den Raumbezug bedeckt sind. Auch für die Modellierung der Geometrie der Raumbezüge durch Rasterzellen wird noch kein Geoinformationssystem benötigt. Es genügt ein relationales Datenbanksystem zur Repräsentation der Raumbezüge. Topologisch/geometrische Beziehungen wie das Enthaltensein oder die Überlappung von Raumbezügen lassen sich durch Abfragen in der Datenbanksprache SQL überprüfen. Als dritte Möglichkeit besteht die Verwendung von Vektorgeometrien, die eine sehr präzise Repräsentation der Geometrie von Raumbezügen ermöglichen. In diesem Fall muß ein echtes Geoinformationssystem genutzt werden. Hierfür kommt sinnvollerweise ein Geoinformationssystem in Frage, das auf einem relationalen Datenbanksystem aufbaut.

Der Gazetteer kann nun auf verschiedene Weisen für die Recherche und Katalogisierung von Informationen in einem Metainformationssystem genutzt werden:

- Mit Hilfe eines *Raumbezugsbrowsers* ist es möglich, Raumbezüge in textueller Form zusammen mit hierarchisch übergeordneten oder untergeordneten sowie überlappenden Raumbezügen darzustellen und zu diesen Raumbezügen zu navigieren. Die Namen der aufgesuchten Raumbezüge können bei der Recherche nach Informationen als Suchkriterien übernommen werden. Sie können auch genutzt werden für die manuelle Zuordnung von Raumbezügen zu Informationsressourcen für deren Katalogisierung im Metainformationssystem.
- Darüber hinaus können die geographischen Namen im Gazetteer auch verwendet werden, um eine *Textanalyse* von zu katalogisierenden Dokumenten hinsichtlich ihres Raumbezugs vorzunehmen. Eine textuelle Suche nach geographischen Namen in einem zu katalogisierenden Dokument liefert die gewünschten Raumbezüge, unter denen das Dokument im Metainformationssystem abgelegt werden soll. Nach diesem Prinzip lassen sich auch Robot-Programme entwickeln, die das Internet nach raumbezogenen Informationen absuchen und eine automatische Georeferenzierung vornehmen.
- Herkömmliche *Suchmaschinen* können mit Hilfe der Information aus dem Gazetteer für die raumbezogene Recherche angepaßt werden. Hierzu ist es möglich, einen Präprozessor für Benutzeranfragen zu entwickeln, der in einer Benutzeranfrage die Namen von Raumbezügen identifiziert und z.B. um weitere Namen von überlappenden Raumbezügen erweitert.
- Die geometrische Repräsentationsform von Raumbezügen ermöglicht auch *graphische Techniken* der Informationsrecherche und -katalogisierung. So lassen sich Raumbezüge durch direkte Manipulation mit Hilfe einer Maus spezifizieren (Abbildung 2). Alternativ ist eine Spezifikation von Raumbezügen durch Angabe von Koordinaten möglich.
- Ergebnisse von Anfragen an ein Metainformationssystem lassen sich auf einer kartographischen Oberfläche visualisieren. So kann eine Liste von gefundenen Informationsressourcen, welche jeweils einen Raumbezug besitzen, in Form sogenannter *Footprints* (Fußabdrücke) auf einer Landkarte dargestellt werden (Abbildung 3). Durch Anklicken eines solchen Footprints kann man dann auf die gewünschte Information direkt zugreifen.

Das hier vorgestellte Konzept, Informationen mit Raumbezügen zu versehen, ermöglicht neue Möglichkeiten für die Katalogisierung und Recherche von Informationen auf dem Internet. Als besonderer Vorteil erweist sich die Möglichkeit, Informationen über ihre Raumbezüge graphisch darzustellen und zu selektieren. Zugleich ist es mit Hilfe von Robot-Programmen



**Abbildung 3:** Darstellung eines Rechercheergebnisses in Form von sogenannten Footprints. Die Raumbezüge der gefundenen Informationen (im gezeigten Beispiel handelt es sich um Luftbilder) sind auf einer Karte dargestellt. Die gezeigte Funktionalität ist Teil der Alexandria Digital Library, die an der University of California at Santa Barbara entwickelt wurde.

und automatischer Textanalyse möglich, die Raumbezüge von Informationsangeboten auf dem Internet automatisch zu bestimmen.

Insgesamt zeigt sich, daß die Georeferenzierung von Informationen auf dem Internet vielfältige Möglichkeiten bietet, um die Informationsflut zu beherrschen, mit der die Informationsgesellschaft zunehmend frei Haus konfrontiert wird.