

Fach-, raum- und zeitbezogene Katalogisierung und Recherche von Umweltinformationen auf dem Internet

Wolf-Fritz Riekert, Klaus Tochtermann, Gerlinde Wiest, Margit Gaul
FAW Ulm

Jürgen Seggelke, Birgit Mohaupt-Jahr
Umweltbundesamt Berlin

Zusammenfassung

Die auf dem Internet verfügbare Information nimmt sowohl in ihrem Umfang als auch in ihrer Qualität unvermindert zu. Um dieses Informationsangebot gezielt recherchieren zu können, etablieren sich derzeit erste Metainformationssysteme, die Informationen über die zu suchenden Informationen, sogenannte Metainformationen, bei der Suche ausnutzen. Diese durchsuchen im Gegensatz zu herkömmlichen Suchmaschinen nicht das Informationsangebot an sich, sondern Metainformationsbestände. Von diesem Ansatz können Benutzer in zweierlei Hinsicht profitieren. So entstehen zum einen neue Anfrageformen, zum anderen verbessert sich die Qualität der Rechercheergebnisse. Dieser Beitrag stellt nun ein Metainformationssystem für Umweltinformationen vor, in dem Fach-, Raum- und Zeitbezüge ausgenutzt werden, um Metainformationsbestände aufzubauen und einfach recherchierbar zu machen. Dabei wird neben einer Darstellung konzeptioneller Aspekte ausführlich der sich bereits im Einsatz befindliche Prototyp beschrieben.

Abstract

The amount and quality of information available on the Internet increases steadily. To enhance the search and retrieval of this information, meta-information systems are being developed. Unlike earlier search engines, their search mechanisms base upon meta-information (i.e., information on the information to be searched for) rather than the information itself. Users can benefit from this approach in two ways: Firstly, they are provided with new types of queries; secondly, the quality of retrieval results is improved. This paper presents a meta-information system which is particularly designed for environmental information. It allows to make use of semantic, spatial, and temporal relationships for the search and retrieval process. Besides concepts, this paper primarily presents a comprehensive description of an already implemented and running prototype.

1 Einleitung

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 fast unüberschaubaren Angebot von Informationsdiensten zu orientieren. Hier sind neuartige Informationssysteme erforderlich, um die derzeitige, sicherlich unbefriedigende Situation zu verbessern.

Neueste Entwicklungen in diesem Bereich bringen nun eine neue Generation von Systemen hervor, die sogenannten HIM-Systeme (Hypertext – Information Retrieval – Multimedia). Ziel von HIM-Systemen ist es, Synergieeffekte zwischen den drei genannten Bereichen auszunutzen und so eine neue Generation elektronischer Informationssysteme zu etablieren. Hiervon verspricht man sich Lösungen, um beispielsweise das unüberschaubare Informationsangebot auf dem Internet Nutzern besser zugänglich zu machen. Der Vollständigkeit halber sei erwähnt, daß natürlich auch HIM-Systeme denkbar sind, die losgelöst vom Internet operieren können. 1995 fand in Konstanz eine erste HIM-Tagung statt [KuRi95], die zum Ziel hatte, Forscher aus den drei Teildisziplinen zusammenzubringen, um die Entwicklung integrierter HIM-Systeme voranzutreiben. Inzwischen, mehr als zwei Jahre später, sind erste solcher Systeme verfügbar. Im Rahmen dieses Beitrages wird mit dem Locator des Umweltinformationsnetzes Deutschland GEIN (German Environmental Information Network) eine Variante eines solchen HIM-Systems beschrieben.

Der GEIN-Prototyp wurde vom FAW Ulm im Auftrag des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit und des Umweltbundesamts in Berlin entwickelt. Im Zusammenhang mit diesem Auftrag wurde GEIN als deutscher Beitrag zu dem G7-Pilotprojekt Environment and Natural Resources Management (ENRM) [ENRM97] auf der G7-Konferenz „Information Society and Development“ in Südafrika im Mai 1996 vorgestellt. Seit Anfang 1997 befindet sich der GEIN-Prototyp im Testeinsatz beim Umweltbundesamt.

Der Beitrag ist nun wie folgt aufgebaut. In Kapitel 2 wird die Notwendigkeit der Entwicklung von Metainformationssystemen mit neuen Qualitäten motiviert. Daran schließt sich die Beschreibung der Katalogobjekte des im Metainformationssystem GEIN verwendeten Katalogs. Wie unter Verwendung dieses Katalogs recherchiert werden kann, wird in Kapitel 4 verdeutlicht. Kapitel 5 geht kurz auf die technische Realisierung ein. Schließlich stellt Kapitel 6 den Bezug zu existierender Literatur her, bevor der Beitrag mit einer Diskussion und einem Ausblick in Kapitel 7 schließt.

2 Aktueller Stand und Motivation

Wie in [Lev95] dargestellt, gewinnt die Katalogisierung von Informationen auf dem Internet in zunehmendem Maße an Bedeutung. Große amerikanische Organisationen, wie das Online Computer Library Center (OCLC), haben bereits mit entsprechenden Projekten auf dieses Bedürfnis reagiert [Wei95, Jul95]. Daß im Bereich Katalogisierung und Metainformationen (unter Metainformationen verstehen wir Informationen über Informationen) auf dem Internet Handlungsbedarf besteht, beweisen auch die derzeit noch sehr beschränkten Möglichkeiten der existierenden Suchhilfen. Hier lassen sich im Internet derzeit drei verschiedene Typen ausmachen: Suchmaschinen, wie AltaVista [AlVi97], systematische Themenkataloge, wie der WWW Virtual Library Subject Catalog [ViLi97], und Metainformationssysteme, wie Galaxy [Gal97] und der Internet Resource Meta Index von NCSA [NCSA97].

Die verschiedenen im Internet angebotenen Suchmaschinen ermöglichen die Suche 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, sogenannte „Robots“, gepflegt wird. Diese verfolgen selbständig das

Geflecht von Verweisen im WWW und nehmen eine Indizierung der gefundenen HTML-Seiten vor. 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 Anwendungsgebiete mit komplexen Beziehungen zwischen den einzelnen Informationseinheiten aus.

Systematische Themenkataloge erlauben eine sehr zielorientierte Suche nach Informationen. Im Gegensatz zu Suchmaschinen werden nicht die zu einer Anfrage passenden Dokumente geliefert. Vielmehr müssen sich Nutzer selbst sehr zielgerichtet durch einen solchen Katalog bewegen. Ein Nachteil dieser Systeme besteht darin, daß Nutzer nicht wissen, welche Informationen sie hinter einem Verweis aus dem Katalog erwartet. So stimmt häufig die Erwartungshaltung der Nutzer nicht mit der Systematik des Katalogs überein. Navigiert man beispielsweise in dem WWW Virtual Library Subject Catalog ausgehend von dem Haupteintrag „Geographie“ zu dem Untereintrag „Zypern“, so würde man Informationen zur Geographie Zyperns erwarten. Tatsächlich findet man jedoch eine Seite mit allgemeinen Informationen über Zypern, wie etwa Tourismus, Sport, Politik etc. Die mangelnde Übereinstimmung zwischen konkreter Systematik und Erwartungshaltung der Benutzer erweist sich als besonders nachteilig, wenn Benutzer Informationen zu einem Thema suchen, das nicht als Hauptkategorie in einem Themenkatalog vorkommt. In diesem Fall bleibt nur die mühsame Suche durch alle vorhandenen Kategorien, in der Hoffnung, das gewünschte Thema als eine Unterkategorie zu finden. Aufgrund einer häufig relativ groben Systematik ist es weiterhin nicht möglich, Informationen nach verschiedenen Kriterien zu suchen. Um beim obigen Beispiel zu bleiben, kann ein Themenkatalog in der Regel nicht gleichzeitig nach mehreren Kriterien „Südfrucht“, „Bodensee“ und einer Zeitangabe, etwa 1990-1996, durchsucht werden. Eventuell vorhandene Artikel über den Anbau von Apfelsinen auf der Insel Mainau im Jahr 1993 können somit auch nicht ohne weiteres gefunden werden.

Metainformationssysteme können im Prinzip genauso eingesetzt werden wie Suchmaschinen. Im Gegensatz zu diesen liegt ihnen jedoch ein Katalog zugrunde, in denen die katalogisierten Ressourcen systematisch indexiert sind. Die Qualität der Suchergebnisse hängt somit stark von der Komplexität des Kataloges ab. Werden dort keine semantischen Informationen über die erfaßten Dokumente abgelegt, können diese auch nicht bei einer Suche ausgenutzt werden. Häufig, wie auch bei Galaxy, kann nur die Systematik des Katalog bei der Suche ausgenutzt werden. Bei der Suche nach „Südfrucht“ und „Bodensee“ stößt man also auf ähnliche Probleme wie bei Suchmaschinen.

Wie die obigen Ausführungen zeigen, sind derzeitige Suchhilfen im Internet nicht für Anwendungsgebiete geeignet, in denen neben komplexen Abhängigkeiten zwischen einzelnen Informationseinheiten auch nicht standardisiertes bzw. nicht standardisierbares Vokabular vorkommt. Hierfür sind neue Arten von Suchhilfen erforderlich. So beziehen sich viele der in Wirtschaft und Verwaltung erhobenen und genutzten Daten auf die natürliche oder technisch geprägte Umwelt des Menschen. Komplexe Zusammenhänge zwischen einzelnen Da-

ten sind hierbei die Regel. Häufig besitzen derartige Umweltdaten Fach-, Raum- und Zeitbezüge, die, falls sie explizit modelliert sind, bei der Recherche ausgenutzt werden könnten. Beispielsweise zielt die im obigen Beispiel angeführte Recherche auf Informationen mit Fachbezug „Südfrucht“, Raumbezug „Bodensee“ und Zeitbezug „1990-1996“ ab.

Seit Mitte der achtziger Jahre ist die Nachfrage nach zuverlässigen und aufgewerteten Daten und Informationen über die Umwelt merklich angestiegen. Aus diesem Grund haben viele deutsche Verwaltungen, Behörden, Forschungsorganisationen und regierungsunabhängige Organisationen mächtige Informationssysteme erstellt, in denen sie Informationen über die Umwelt sammeln, verwalten, analysieren und vorstellen. In jüngster Zeit werden vermehrt Daten, Dienste und Dokumente dieser Umweltinformationssysteme weitläufig über das Internet bereitgestellt. Eine wachsende Zahl von WWW-Servern machen Informationen zur Umwelt für Benutzer aus den Bereichen Politik, Verwaltung, Wirtschaft, Forschung und der Öffentlichkeit zugänglich. Allerdings sind die bereitgestellten Informationsbestände angesichts ihres Umfangs und ihrer Komplexität mit den herkömmlichen, oben beschriebenen Ansätzen schwierig und oft nur unvollständig zu recherchieren.

1995 begann der Bund-/Länder-Arbeitskreis Umweltinformationssysteme damit, das Umweltinformationsnetz Deutschland (GEIN) zu konzipieren, das die vorhandenen einschlägigen Server integriert. Aufgabe von GEIN ist es, einen Rahmen für alle nationalen und internationalen Umweltinformationsserver zu schaffen, die für Deutschland und über seine Grenzen hinaus relevant sind [SeLe96, Rie97a]. Als Kern von GEIN wurde ein Metainformationssystem, der sogenannten GEIN-Locator entwickelt. Gegenüber herkömmlichen Ansätzen wurden die Möglichkeiten des GEIN-Locators zur Katalogisierung und Recherche von Informationen um besondere semantische Kriterien entscheidend erweitert. Insbesondere werden drei Arten von Beschreibungselementen für Informationen unterschieden, nämlich Fachbezüge, Raumbezüge und Zeitbezüge. Ein besonderes Merkmal von GEIN liegt darin, daß in einer Recherche gleichzeitig alle Bezugsformen oder aber nur ausgewählte verwendet werden können. Auch wenn die in GEIN vorhandenen Informationsressourcen sich grundsätzlich auf die Umweltthematik beziehen, ist der GEIN-Locator so allgemein konzipiert, daß auch andere raum- und zeitbezogene Informationsbestände, etwa aus den Gebieten Regionalplanung, Verkehr, Landwirtschaft oder Statistik, damit katalogisiert und recherchiert werden können. Derzeit umfaßt der Katalog von GEIN ca. 2000 Einträge.

3 Die Katalogobjekte

Die Ressourcen, die in dem GEIN zugrundeliegenden Katalog nachgewiesen werden, stammen zu einem Teil aus dem Internet, zu einem anderen Teil aber auch aus Ressourcen eines behördeneigenen Intranets. Der GEIN zugrundeliegende Katalog ist eine Sammlung von Katalogobjekten, die über sogenannte Thesaurusbegriffe für den Fachbezug, Gazetteerobjekte für den Raumbezug und Zeitintervalle für den Zeitbezug beschrieben sind. Ein Katalogobjekt hat darüber hinaus einen Namen, eine URL, d.h. eine Netzwerkadresse, unter der es im Internet oder einem Intranet zugänglich ist, sowie weitere Attribute, in welchen unter anderem Informationen über die datenhaltende Stelle kodiert sind. Weitere Informationen zum strukturellen Aufbau des Katalogs sind in [ToRi97] zu finden.

Die Deskribierung der zu katalogisierenden Informationsressourcen wird in der Regel computerunterstützt von den Informationsanbietern durchgeführt [ToRi97]. Weiterhin können Inhalte anderer Metainformationssysteme, etwa des Umweltdatenkatalogs (UDK) [GüLe96], importiert und an die Struktur des Katalogs in GEIN angepaßt werden.

3.1 Datenbankabfragen als Katalogobjekte

Ein großes Problem besteht häufig darin, Daten aus vorhandenen Datenbanken durch einen Katalog recherchierbar zu machen. Aufgrund des riesigen Datenvolumens ist es jedoch nicht möglich, jede Datenbank zu indexieren und zu katalogisieren. Vor diesem Hintergrund werden bei GEIN typische Benutzeranfragen identifiziert und als Katalogobjekt abgelegt. Der Katalog weist damit vorhandene Daten nicht nur direkt nach, sondern auch indirekt über Dienste, durch die die Daten erzeugt werden können.

3.2 Der Thesaurus

Im Katalog von GEIN dient der Umweltthesaurus des Umweltbundesamtes [Bat94] als Verwaltungsstruktur für Fachbezüge. Katalogobjekte lassen sich mit den Begriffen aus dem Thesaurus über eine n:m-Beziehung miteinander verknüpfen. Der Thesaurus besitzt zweierlei Funktionen: Zum einen stellen die im Thesaurus enthaltenen Begriffe ein Vokabular an Schlagworten dar, das für die Verschlagwortung und die Recherche von Umweltinformationen in GEIN genutzt werden kann. Zum andern behandelt der Thesaurus diese Begriffe als Entitäten, die über Beziehungen zu einem semantischen Netz verknüpft sind. Im wesentlichen handelt es sich dabei um drei Beziehungstypen, nämlich die *Ober-/Unterbegriff-Beziehung*, die Beziehung zwischen *verwandten Begriffen* und die *Synonymbeziehung*. Mit Hilfe dieser Beziehungstypen läßt sich in einem solchen Thesaurus z.B. festlegen, daß „Südfrucht“ ein Oberbegriff des Begriffs „Orange“ und dieser wiederum ein Synonym des Begriffs „Apfelsine“ ist.

Um auch die Informationssuche in einem internationalen Rahmen durch den Thesaurus verbessern zu können, ist dieser zudem um eine englischsprachige, von der Anlage her mehrsprachige Komponente des Umweltbundesamtes ergänzt.

3.3 Der Gazetteer

Ein Gazetteer (englischer Begriff für „Ortsverzeichnis“) ist eine Verwaltungsstruktur für Raumbezüge und kann ähnlich wie ein Thesaurus verstanden werden. Die Elemente des Gazetteers, die sogenannten Gazetteerobjekte, stehen für geographische Orte oder Gebiete. Der Raumbezug von Ressourcen ist repräsentiert durch eine n:m-Beziehung zwischen Katalogobjekten und Gazetteerobjekten. Die Möglichkeit, Katalogobjekte mit mehreren Gazetteerobjekten in Beziehung zu bringen, ist von großer Bedeutung, da sich Informationsressourcen im Umweltbereich oft auf mehrere geographische Gebiete beziehen. So kann eine Umweltvorschrift verschiedene Raumbezüge haben, etwa dann, wenn in dieser Vorschrift mehrere geographische Namen, z.B. Gemeinidenamen, vorkommen. Genauso ist es

umgekehrt erforderlich, daß ein Gazetteerobjekt mehrere Katalogobjekte beschreiben kann. Beispielsweise können zu einem Naturschutzgebiet mehrere Forschungsberichte existieren.

Ein einem Katalogobjekt zuordenbares Gazetteerobjekt hat als Attribute einen geographischen Namen sowie eine in der Regel zweidimensionale Geometrie. Somit ist es möglich, Raumbezüge sowohl bei der Katalogisierung wie auch bei der Recherche von Informationsressourcen auf zweierlei Weisen zu referenzieren. Der geographische Name dient zur textuellen Bezeichnung eines Raumbezugs einer Information (z.B. der Name Mainau in dem in eingangs gegebenen Beispiel). Über die Geometrie wird durch eine Menge geographischer Koordinaten die Lage und Ausdehnung des Raumbezugs festgelegt.

Ähnlich wie Thesaurusbegriffe können auch Gazetteerobjekte untereinander durch Beziehungen verknüpft werden. So ist es häufig erforderlich und sinnvoll, Gazetteerobjekte hierarchisch zu ordnen. Beispielsweise ist bei Verwaltungseinheiten die Hierarchie Bundesland, Landkreis, Gemeinden sehr geläufig. Analog sind Postleitgebiete entsprechend ihrer dezimalen Gliederung hierarchisch geordnet.

Von besonderer Bedeutung sind in GEIN die geometrisch/topologischen Beziehungen von Raumbezügen, wie Überlappung, Ineinanderenthaltensein und gegenseitige Berührung. Diese Beziehungen sind derzeit jedoch nicht explizit im Katalog von GEIN verankert. Hierfür gibt es zwei Gründe: 1) die Beziehungen können aus der Geometrie automatisch ermittelt werden; 2) da z.T. Katalogeinträge von Hand erstellt werden, müßten Nutzer alle bereits vorhandenen Gazetteerobjekte darauf prüfen, ob eine der Beziehungen zu dem neu angelegten Gazetteerobjekt zutrifft. Dies könnte man zwar auch automatisieren, allerdings entstünden daraus sehr große Tabellen mit redundanten Inhalten, was eine größere Komplexität und damit Fehlerträchtigkeit in sich birgt.

Beziehungen zwischen Gazetteerobjekten erleichtern die Recherche nach Informationen auf beträchtliche Weise. Wenn die hierarchischen bzw. topologischen Beziehungen im Gazetteer korrekt repräsentiert sind, brauchen bei der Recherche nach Informationsressourcen nicht exakt dieselben Raumbezüge angegeben zu werden, die bei deren Katalogisierung verwendet wurden. Eine geeignet gestaltete Recherchesoftware, die die Beziehungen zwischen Gazetteerobjekten sinnvoll auswertet, ist imstande, topologisch äquivalente oder verwandte Raumbezüge zu erkennen und die zugehörigen Katalogobjekte zu finden.

Die Geometrie eines Gazetteerobjekts kann auf drei unterschiedliche Weisen modelliert werden.

1. **Umschließendes Rechteck.** Die einfachste Möglichkeit ist die Repräsentation der Geometrie in Form eines umschließenden Rechtecks. Dies bewirkt eine starke Vergrößerung der Geometrie von Raumbezügen. Soll z.B. eine Stadt als Gazetteerobjekt abgelegt werden, so ist es immer notwendig, ein umschließendes Rechteck anzugeben, welches die Stadt in seiner maximalen Ausdehnung beschreibt. Dies führt dazu, daß Randzonen, die nicht mehr zum Stadtgebiet gehören, im Raumbezug eingeschlossen sind. Dieses Problem verschärft sich noch, wenn Geobjekte mit linearer Ausdehnung wie z.B. Straßen und Flüsse als Raumbezüge genutzt werden sollen. Allerdings hat dieser Ansatz den Vorteil, daß für die Repräsentation des Gazetteers kein echtes Geoinformationssystem benötigt wird.

2. **Rasterzellen.** Eine detailfreudigere Modellierung der Geometrien von Gazetteerobjekten kann mit Hilfe von Rasterzellen erreicht werden. Hier wird das Plangebiet in eine hinreichend feine Rasterung von Planquadraten zerlegt. Diese Planquadrate lassen sich durchnumerieren und daher durch eine einfache Zahl identifizieren. Die Geometrie eines Raumbezugs ist gegeben durch die Menge aller Planquadrate, die durch den Raumbezug bedeckt sind. Allerdings muß man sich bei der Wahl der Längen- und Breitenausdehnung auf ein im Hinblick auf die Menge der abzuspeichernden Daten auf ein handhabbares Maß festlegen. Wird z.B. die Größe der Rasterzellen auf je ein Kilometer Längen- und Breitenausdehnung festgelegt, so ergeben sich bei einem aufzurasternden Gebiet mit einer Ausdehnung von 200 mal 200 Kilometern bereits 40.000 Rasterzellen. Auch das oben angesprochene Problem der genauen Abbildung der Gebietsgeometrien ist mit der Einführung von Rasterzellen noch nicht gänzlich behoben, jedoch in weitaus geringerem Ausmaß vorhanden. Die Genauigkeit ist dabei umgekehrt proportional zur Dimensionierung der Rasterzellen. 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, die Berührung oder die Überlappung von Raumbezügen lassen sich durch Abfragen in der Datenbanksprache SQL überprüfen.
3. **Vektorgeometrien.** Als dritte Möglichkeit besteht die Verwendung von Vektorgeometrien, die eine sehr präzise Repräsentation der Geometrie von Raumbezügen ermöglichen, insbesondere bei Geoobjekten mit linearer Ausdehnung. In diesem Fall muß ein echtes Geoinformationssystem genutzt werden. Hierfür kommt sinnvollerweise ein Geoinformationssystem in Frage, das auf einem relationalen Datenbanksystem aufbaut.

In einem ersten Prototypen von GEIN werden Raumbezüge mittels umschließender Rechtecke realisiert. Um einen höheren Detaillierungsgrad zu erreichen, wird dies jedoch derzeit durch die Verwendung von Rasterzellen mit einer Ausdehnung von 1 mal 1 Kilometer abgelöst.

3.4 Zeitinformation

Schließlich besteht zu jedem Katalogobjekt eine n:m-Beziehung zu Zeitintervallen, die für die Auswertung von Zeitbezügen verwendet werden. Mehrere Zeitintervalle pro Katalogobjekt sind erforderlich, da z.B. bei Umweltgesetzen mit dem Zeitpunkt des Inkrafttretens und dem Zeitraum der Verabschiedung mehrere zeitbezogene Informationen vorliegen können.

4 Recherche mittels GEIN

Mittels GEIN können die im Katalog verankerten Bezüge bei einer Recherche ausgenutzt werden. Ein wichtiges Merkmal ist dabei, daß Nutzer entscheiden können, ob sie eine, zwei oder alle drei Bezugsformen nutzen möchten. Diese Freiheit ermöglicht ihnen z.B. eine rein fachbezogene Recherche ohne Berücksichtigung von Raum- und Zeitbezügen.

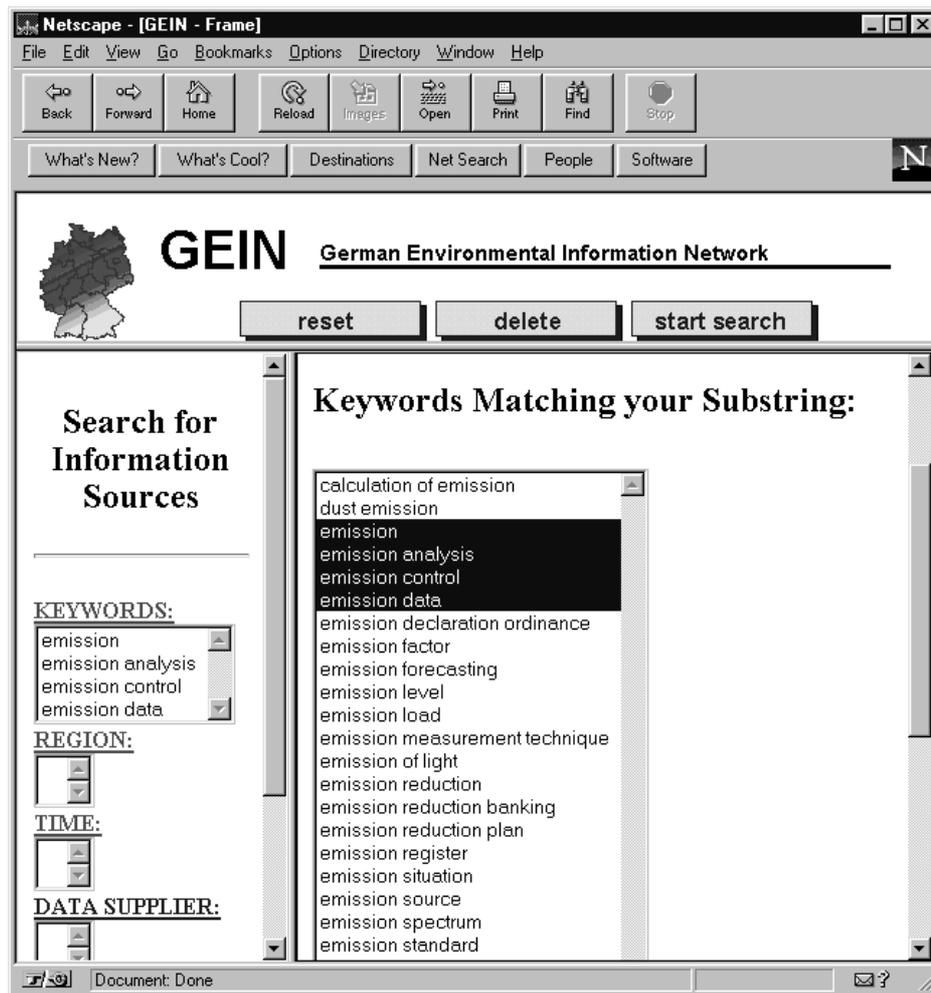


Abb. 1: Auswahl von Fachbezügen in GEIN

Abbildung 1 stellt die Rechercheoberfläche von GEIN dar. Auf der linken Seite ist unter den Überschriften „Keywords“ („Schlagwort“ in der deutschen Variante des Systems), „Region“ und „Time“ („Zeit“) die aktuelle Recherche mit den spezifizierten Fachbezügen, Raumbezügen und Zeitbezügen dargestellt. Im folgenden wird dargestellt, wie die verschiedenen Bezugsformen für eine Suche festgelegt werden können. Um den Lesern ein Gefühl dafür zu geben, wie Benutzer über eine intuitive und hypermediale Benutzungsschnittstelle einen Katalog mit den genannten Bezugsformen für die Recherche einsetzen können, werden die jeweiligen Möglichkeiten anhand entsprechender Bildschirmabzüge von GEIN illustriert. Die Ergebnisdarstellung einer Recherche wird in Abbildung 3 am Beispiel der deutschen Benutzeroberfläche des Systems gezeigt.

4.1 Festlegen des Fachbezugs

Häufig kennen Nutzer nicht alle im Thesaurus verwendeten Deskriptoren. Daher kann zum Festlegen des Fachbezugs eine Teilwortsuche angestoßen werden. Das Ergebnis dieser Su-

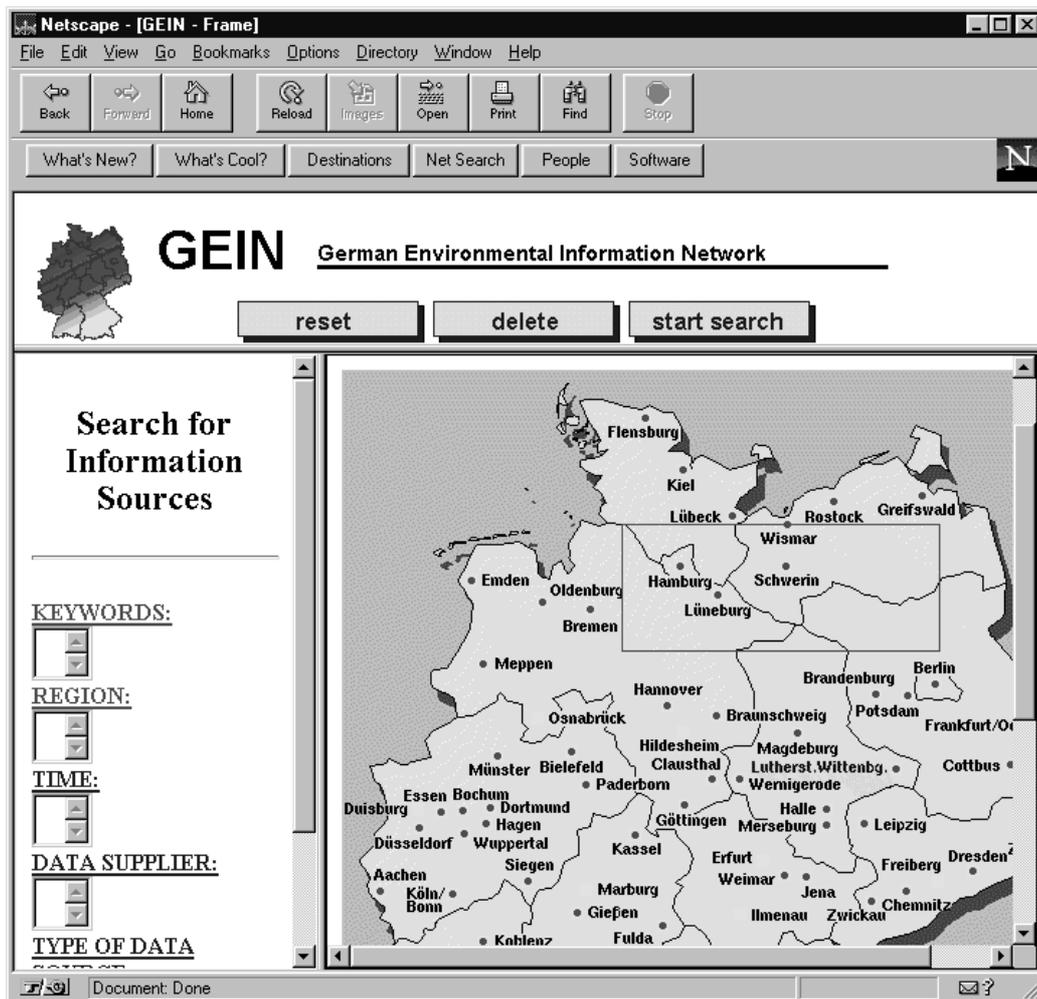


Abb. 2: Interaktive Selektion von Raumbezügen in GEIN

che ist eine Liste von Deskriptoren, die das Teilwort enthalten. Abbildung 1 zeigt z.B. einen Teil der Liste von Deskriptoren, die für das Teilwort 'Emission' gefunden wurde. Aus dieser Liste können Nutzer beliebig viele Deskriptoren für den Fachbezug einer Recherche auswählen. Für die gefundenen Deskriptoren kann sich der Nutzer auch die im Thesaurus vorhandenen Ober-, Unterbegriffe, verwandte Begriffe und Synonyme anzeigen lassen und gegebenenfalls dort weitere Fachbezüge für die Recherche auswählen (dieser Zusammenhang ist in Abbildung 1 nicht ersichtlich).

4.2 Festlegen des Raumbezugs

Da im Gazetteer sowohl geographische Namen also auch Geometrien für Raumbezüge verwaltet werden, kann der Gazetteer auf zwei verschiedene Weisen für die Recherche von Informationen genutzt werden (eine davon ist in Abbildung 2 gezeigt):

1. 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 für die manuelle Zuordnung von Raumbezügen zu Informationsressourcen für deren Katalogisierung im Metainformationssystem genutzt werden.

2. Die geometrische Repräsentationsform von Raumbezügen ermöglicht auch interaktive Selektion von Raumbezügen. So lassen sich Raumbezüge durch direkte Manipulation mit Hilfe einer Maus spezifizieren (vgl. Abbildung 2). Alternativ ist eine Spezifikation von Raumbezügen durch Angabe von Koordinaten möglich. Ähnlich wie beim Thesaurus, wo für gefundene Deskriptoren Synonyme etc. angeschaut und ausgewählt werden konnten, ist es beim Gazetteer möglich, sich für einen selektierten Raumbezug die Gebietshierarchie anzeigen zu lassen, um dann dort weitere Geometrien für die Recherche auszuwählen (dieser Sachverhalt ist in Abbildung 2 nicht ersichtlich).

4.3 Festlegen weiterer Suchkriterien

Hier soll noch kurz auf Zeitbezüge, auf datenhaltende Stellen und auf die Art der Informationsressource eingegangen werden. Bei der Recherche über Zeitbezüge können entweder Zeitpunkte oder aber ganze Zeitintervalle über ein entsprechendes Formular ausgewählt werden. Wird ein Zeitpunkt angegeben, so werden bei der Recherche alle diesen Zeitpunkt enthaltenden Zeitintervalle berücksichtigt. Bei Zeitintervallen werden alle Zeiträume berücksichtigt, die den angegebenen Zeitbezug überschneiden. Zudem kann man den Suchraum auf bestimmte datenhaltende Stellen und auf die Art der Informationsressource (Bericht, Dienst, Ressource des Umweltdatenkatalogs) einschränken. Derzeit sind 18 datenhaltende Stellen an GEIN angeschlossen, darunter das Umweltbundesamt, die Umweltministerien in Baden-Württemberg, Bayern, Niedersachsen u.v.a.

4.4 Beispiel einer Suchanfrage und zugehörigem Suchergebnis

In diesem Abschnitt wird vorgestellt, wie das Ergebnis einer Anfrage in GEIN präsentiert wird.

Die Abbildung 3 zeigt in der deutsche Benutzungsschnittstelle von GEIN im linken Recherchefenster eine Anfrage und im rechten Recherchefenster das zugehörige Suchergebnis. Mit der Anfrage wird nach Informationen über Emissionen (Fachbezug) aus der Region Baden-Württemberg (Raumbezug) im Jahr 1996 (Zeitbezug) gesucht. Datenhaltende Stelle und Art der Informationsressource sind nicht eingeschränkt. Rechts sind die im Katalog gefundenen Katalogeinträge dargestellt. Die Information umfaßt wenn möglich einen Verweis auf ein anderes an das Internet angebundene Informationssystem, Informationen über die datenhaltende Stelle sowie Informationen über den Ansprechpartner. Durch Folgen der Verweise können die Informationen abgerufen werden. So kann im gezeigten Beispiel über den Verweis „UFIS-Selektor Kraftwerke“ zu dem Umweltführungsinformationssystem (UFIS) [Rie97b] des Landes Baden-Württemberg und damit von den Metainformationen zu den eigentlichen Informationen gewechselt werden.



Abb. 3: Beispiel einer Anfrage und zugehörigem Suchergebnis

5. Technische Realisierung

GEIN wird auf der Basis des WWW mit den dort verfügbaren Mitteln entwickelt. So kommen etwa HTML-Formulare, PERL-Skripte und Java zum Einsatz. Als Frontend können beliebige Webbrowser eingesetzt werden. Als Server wird auf eine Variante von Netscape zurückgegriffen. Alle Katalogobjekte sind nun in einer Oracle-Datenbank abgelegt. Für deren Anbindung ans WWW sowie für die Anbindung von Datenbanken mit im Katalog nachgewiesenen Beständen wurde vom FAW im Rahmen des Umweltinformationssystems Baden-Württemberg ein generisches Umweltdatenbank-Gateway entwickelt, das in GEIN auch genutzt wird [Rie97b]. Dieses Umweltdatenbank-Gateway umfaßt einen Interpreter namens WebQuery zur Verarbeitung von Nutzeranfragen [Gau95]. Da dies eine Besonderheit der realisierten technischen Lösung ist, soll es näher beschrieben werden.

Wenn genau eine Datenbank zum Zugriff über das WWW anzubinden ist, genügt in der Regel eine speziell auf diese Datenbank abgestimmte Lösung. Werden jedoch in einem Katalog Ressourcen von beliebig vielen Datenbanken in Form von Anfragen, die diese Res-

sources erzeugen, abgelegt, ist dieser wenig flexible Weg nicht mehr gangbar. Der Grund liegt darin, daß für jede Anfrage eine speziell abgestimmte Lösung fest zu implementieren wäre. Dies ist mit Inflexibilität und hohem Aufwand verbunden. Vielmehr bietet sich die Entwicklung eines generischen Umweltdatenbank-Gateways, wie WebQuery, an, das es erlaubt, beliebige relationale Datenbanken an das WWW anzuschließen. WebQuery stellt dabei unabhängig von der angeschlossenen Datenbank allgemeine Funktionen hinsichtlich Mehrbenutzerbetrieb, Benutzerverwaltung, Statistik, Download etc. bereit. Eine besondere Komponente ist die Sitzungsverwaltung. Bekannterweise ist HTTP zustandlos, so daß Verbindungen zwischen Klienten und Server sofort nach Antwort des Servers auf eine Klientenanforderung abgebrochen werden. Geschachtelte Transaktionen, d.h. solche, die auf einem Zwischenergebnis einer zuvor abgeschickten Transaktion operieren, sind nicht ohne weiteres möglich. Um dies dennoch zu ermöglichen, wird ein Sitzungsmanager angeboten, der für die Dauer einer Sitzung Kontexte aufrecht erhält. Unter einer Sitzung wird dabei eine zeitliche Abfolge von Dienstanforderungen verstanden, die zwischen einem Sitzungsstart und dem Sitzungsende von einem Benutzer eines Klientensystems an ein Serversystem gerichtet werden.

WebQuery besteht aus drei Komponenten, die serverseitig laufen. Die erste Komponente generiert aus den Eingaben, die über der Rechercheoberfläche von GEIN festgelegt wurden, ein SQL-Statement. Dieses wird von einer zweiten Komponente, der Verbindungskomponente, an die gewünschte Datenbank abgesetzt. Die Verbindungskomponente empfängt auch das von der Datenbank gelieferte Ergebnis und legt es in einer Datei ab. Auf die Ergebnisdatei greift als dritte Komponente ein Aufbereitungswerkzeug zu und setzt die Daten in HTML um, so daß sie und zurück an die Rechercheoberfläche geliefert und dort dargestellt werden können. Mit WebQuery können nicht nur Selektionen für eine Datenbankanfrage, sondern auch Insert-, Delete- und Update-Statements formuliert werden. Damit ist ein interaktives Arbeiten mit angeschlossenen Datenbanken über einen WWW-Klienten mittels WebQuery möglich.

Diese Dreiteilung des Umweltdatenbank-Gateways hat den Vorteil, daß jede einzelne Komponente ausgetauscht werden kann, falls andere Randbedingungen zu berücksichtigen sind. Weitere Informationen zur technischen Realisierung können in [ToRi97] nachgelesen werden.

6 Literaturdiskussion

Die Mission des U.S. Geological Survey (USGS), einer Einrichtung des amerikanischen Innenministeriums, ist es, geologische, topologische und hydrologische Informationen anzubieten, die zu einer sinnvollen Nutzung der natürlichen Ressourcen beitragen [USGS97]. Eine Komponente von USGS ist das Metainformationssystem National Geospatial Data Clearinghouse [NGDC97]. Dieses bietet insbesondere eine Möglichkeit, um Informationen über Geodaten im USGS zu suchen. Ähnlich wie auch in GEIN können hierfür Raum- und Zeitbezüge, nicht jedoch Fachbezüge ausgenutzt werden.

In der Alexandria Digital Library [Ale97, Smi96] können Ergebnisse von Anfragen an ein Metainformationssystem auf einer kartographischen Oberfläche visualisiert werden. So kann

eine Liste von gefundenen Informationsressourcen, welche jeweils einen Raumbezug besitzen, in Form sogenannter Footprints auf einer Landkarte dargestellt werden. Durch Anklicken eines solchen Footprints kann man dann auf die gewünschte Information direkt zugreifen. Eine solche graphische Darstellung der Anfrageergebnisse ist mit GEIN derzeit nicht möglich.

Für Metainformationssysteme mit Fach-, Raum-, und Zeitbezug sind auch die neuesten Entwicklungen im Bereich Z39.50 [Z39.50] erwähnenswert. Das ursprünglich nur für die Vernetzung von Bibliothekssystemen gedachte Protokoll findet in jüngster Zeit auch immer mehr Verbreitung in anderen Bereichen. Für spezielle Anwendungsbereiche können auf der Basis von Z39.50 sogenannte Profile entwickelt werden, die eine Recherche unter Verwendung anwendungsspezifischer Suchkriterien unterstützen. Zu den für den Umweltbereich relevanten Profilen zählen die in den Systemen GILS (Government Information Locator Service) [GILS97, Chr96] und GELOS (Global Environmental Information Locator Service) [GELOS97] genutzten Profile. GILS ist ein Netz von Metainformationssystemen in den USA, dessen Ziel es ist, der amerikanischen Öffentlichkeit Informationsmaterial (Wirtschaftsdaten, Umweltdaten etc.), das die Regierung mit einem jährlichen Aufwand von mehreren Millionen Dollar erstellt, auf einfache Weise über das Internet zugänglich zu machen. Das zugehörige GILS-Profil enthält Attribute, mit deren Hilfe Anfragen mit Raum- und Zeitbezug übertragen werden können. Fachbezüge können über das Attribut „Cross Reference“ explizit angegeben werden. Auch wenn von Seiten des Protokolls alle relevanten Bezugsformen grundsätzlich unterstützt werden, lassen sich diese in den verschiedenen Metainformationssystemen innerhalb des GILS-Netzwerkes jedoch nur selten finden. Das GELOS-Profil ist relativ neu und wurde im Rahmen des bereits angeführten G7-Pilotprojekts ENRM entwickelt. Es ähnelt stark dem GILS-Profil, besitzt aber Erweiterungen insbesondere in Bezug auf die multilinguale Informationsrecherche.

Das in Entwicklung befindliche Verweis- und Kommunikationssystem Umwelt des Umweltbundesamtes (VKS-Umwelt) knüpft an die Erfahrungen mit GEIN an und bezweckt ebenfalls die Recherche nach diversen Informationsressourcen mit Umweltbezug [SeMo97]. Die Konzeption von VKS-Umwelt sieht jedoch keinen eigenen Katalog vor, vielmehr reicht ein sogenannter „Broker“ die Anfragen an verschiedene Dokumentationsdatenbanken, Metainformationssysteme und Datenkataloge weiter, wobei das Z39.50-Protokoll und weitere Z39.50-Profile genutzt werden sollen. Anders als GEIN kann VKS-Umwelt somit als erster Schritt in Richtung eines Multikatalogsystems verstanden werden (vgl. auch Abschnitt 7).

7 Diskussion und Ausblick

Die eingangs erwähnten Suchmaschinen haben den Vorteil, daß sie potentiell das gesamte WWW durchsuchen können. Nutzer haben dadurch den Vorteil, daß auf eine Anfrage eine relativ große Ergebnismenge zu erwarten ist. Allerdings ist die Qualität der Ergebnisse relativ gering, da nur eine Volltextsuche eingesetzt wird. Im Gegensatz hierzu kann mit GEIN nur in einem zuvor deskribierten Bestand gesucht werden; dort allerdings unter Verwendung von Zeit-, Raum- und Fachbezügen. Da bei der Suche jedoch semantische Kriterien verwendet werden können, sind die Ergebnisse in der Regel von höherer Qualität.

Ein Nachteil von Metainformationssystemen im Internet, aber auch von Suchmaschinen und systematischen Katalogen, liegt sicherlich darin, daß sie auf Internetressourcen verweisen, die nicht unter ihrer Kontrolle liegen. So finden sich in den im WWW verfügbaren Metainformationssystemen immer wieder Katalogobjekte, die auf nicht mehr existierende oder verlegte Internetressourcen verweisen. Um diesem Problem zu begegnen, sind zwei Wege denkbar. Zum einen können Werkzeuge in das System eingebaut werden, die mit einer gewissen Regelmäßigkeit die katalogisierten Internetressourcen auf ihre Gültigkeit hin prüfen. Die andere Möglichkeit besteht darin, die Adressierung von Internetressourcen in Zukunft nicht mehr über absolute URLs, sondern über relative URNs (Uniform Resource Names) zu adressieren. Erste Ansätze hierfür existieren bereits [Toc96]. Für Metainformationssysteme sind URNs jedoch nur hilfreich, um auf einen anderen Server verlegte Internetressourcen weiterhin korrekt nachhalten zu können. Um auch beim Löschen von Internetressourcen Konsistenz in Metainformationssystem wahren zu können, sind Ansätze, wie sie aus Hyper-G/HyperWave bekannt sind, von großem Interesse. Da hier Verweise als eigenständige Objekte und nicht als Bestandteil eines HTML-Dokumentes verwaltet werden, können diese darüber informiert werden, wenn Ziel- oder Ausgangsknoten gelöscht oder verlegt werden. Damit läßt sich zumindest innerhalb und zwischen Hyper-G/HyperWave-Servern Konsistenz wahren [Mau96].

Für die Zukunft bilden Multikatalogsysteme ein interessantes Forschungsthema für uns, das wir insbesondere in dem bereits genannten Projekt VKS-Umwelt verfolgen. Die Idee besteht darin, verschiedene spezialisierte Katalogsysteme zu integrieren, anstatt einen universellen Katalog aufzubauen, in dem so viele Informationsressourcen wie möglich verzeichnet sind. Die verschiedenen Katalogsysteme sollten dabei untereinander verbunden sein, so daß Anfragen gleichzeitig in allen Katalogen verarbeitet werden können. Ein Ergebnis unserer bisherigen Untersuchungen ist, daß hierfür eine „Brokerkomponente“ erforderlich ist, die zwischen der Benutzungsschnittstelle und den einzelnen Katalogsystemen vermittelt. Der Broker übersetzt einheitliche Suchanfragen in die Abfragesprachen aller angeschlossenen Katalogsysteme. Umgekehrt vereinigt der Broker die Rechercheergebnisse der angeschlossenen Katalogsysteme, bevor sie an die Benutzungsoberfläche weitergegeben werden. Der Vorteil hierbei liegt für Nutzer darin, daß sie mit einer einheitlichen Schnittstelle und einer einheitlichen Anfragesprache ihre Recherche durchführen können.

Nach unseren Vorüberlegungen zu diesem Thema scheinen für derartige Anwendungen das Z39.50-Protokoll und seine speziellen Profile, insbesondere GILS und GELOS geeignet zu sein. Der Broker würde hierbei als Z39.50-Client fungieren und die angeschlossenen Katalogsysteme als Z39.50-Server. Der Vorteil dieses Ansatzes läge darin, daß ein derartiges strukturiertes Protokoll genau die Attribute festlegt, die in den angebotenen Katalogsystemen recherchiert werden können. Dies würde z.B. das Problem überwinden, daß in den Katalogsystemen für semantisch gleiche Attribute unterschiedliche Namen verwendet werden (z.B. „Time Period“ und „Temporal Coverage“ als Attributnamen für Zeitangaben.). Zudem könnte durch Abgleich zwischen Broker und Server frühzeitig geklärt werden, nach welchen Attributen in den jeweiligen Katalogsystemen gesucht werden kann.

Die bislang entwickelten und in diesem Beitrag vorgestellten Konzepte ermöglichen neue Formen von Metainformationssystemen im Internet. Beispielsweise können die geographischen Namen im Gazetteer auch verwendet werden, um eine Textanalyse von zu katalogi-

sierenden 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. Darüber hinaus können herkömmliche Suchmaschinen 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. Ähnliche Ansätze sind auch für die Ermittlung von Zeit- und Fachbezügen denkbar.

Das hier vorgestellte Konzept, Informationen mit Raum-, Fach- und Zeitbezügen 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 und automatischer Textanalyse möglich, die Raumbezüge von Informationsangeboten auf dem Internet automatisch zu bestimmen. Insgesamt zeigt sich, daß die auf strukturierten Metainformationen basierende Katalogisierung und Recherche von Informationsressourcen auf dem Internet vielfältige Möglichkeiten bietet, um die Informationsflut zu beherrschen, mit der die Informationsgesellschaft zunehmend frei Haus konfrontiert wird.

Danksagung

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.

Literatur

[Ale97] <http://alexandria.sdc.ucsb.edu/>.

[AIVi97] <http://altavista.digital.com/>.

[Bat94] Batschi, W.D.; Environmental Thesaurus and Classification of the Umweltbundesamt (Federal Environmental Agency) Berlin; Environmental Knowledge Organization and Information Management (Hrsg. P. Stancikova, I. Dahlberg), INDEKS Verlag, Frankfurt/Main, 1994.

[Chr96] Christian, E.J.; GILS - What is it? Where's it going?; D-LIB The Magazine of Digital Library Research, Dezember 1996, <http://www.dlib.org/dlib/december96/12christian.html>.

[ENRM97] <http://www.ispo.cec.be/g7/projects/theme6.html>.

[Gau95] Gaul, M., Kadric, L., Klügl, G., Krist, Th., Riekert, W.-F., Wiest, G.; Management verteilter Umweltinformationsquellen; Projekt Globus – Konzeption und prototypische Realisierung einer aktiven Auskunftskomponente für globale Umwelt-Sachdaten im Umweltinformationssystem Baden-Württemberg Phase II, (Hrsg. Mayer-Föll, R., Jaeschke, A.); Forschungszentrum Karlsruhe, Wissenschaftliche Berichte FZKA 5700, 1995; <http://www.iai.fzk.de/~weidemann/doc/globus2/index.html>.

[Gal97] <http://galaxy.einet.net/galaxy.html>.

Fach-, raum- und zeitbezogene Katalogisierung und Recherche von Umweltinformationen

[GELOS97] <http://ceo.gelos.org/>.

[GILS97] <http://info.er.usgs.gov/gils/gils1p.html>.

[GüLe96] Günther, O., Lessing, H., Swoboda, W.; UDK: A European Environmental Data Catalogue; Proc. Third International Conference in Integrating GIS and Environmental Modelling. National Center for Geographic Information and Analysis, Santa Barbara, 1996; http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/guenther_oliver/my_paper.html.

[Jul95] Jul, E.; OCLC Internet Cataloging Project; D-LIB The Magazine of Digital Library Research, Dezember 1995, <http://www.dlib.org/dlib/december95/briefings/12oclc.html>.

[KuRi95] Kuhlen, R., Rittberger, M.; Tagungsband Hypertext – Information Retrieval – Multimedia, Universitätsverlag Konstanz Bd. 20, 1995.

[Lev95] Levy, D.; Cataloging in the Digital Order; Proceedings of the 2nd Annual Conference on the Theory and Practice of Digital Libraries, Austin, Texas (USA), 31-37, 1995.

[MaJa96] Mayer-Föll, R., Jaeschke, A.; Projekt Globus – Konsolidierung der neuen Systemarchitektur und Entwicklung erster Produktionssysteme für globale Umweltsachdaten im Umweltinformationssystem Baden-Württemberg Phase III, Forschungszentrum Karlsruhe, Wissenschaftliche Berichte FZKA 5900, 1996; <http://www.iai.fzk.de/~weidemann/doc/globus2/index.html>.

[Mau96] Maurer, H.; HyperWave – The Next Generation Web Solution. Addison Wesley, 1996; <http://hyperg.iicm.tu-graz.ac.at/hgbook>.

[NGDC97] <http://nsdi.usgs.gov/nsdi/>.

[NCSA97] <http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/MetaIndex.html>.

[Rie97a] Riekert, W.-F., Wiest, G., Gaul, M., Müntz, B., Henning, I.; Management of Distributed and Heterogeneous Information Resources for Environmental Administrations; Environmental Software Systems, Volume 2 (eds. Denzer, R., Swayne, D.A., Schimak, G.); Chapman & Hall, London, 1997.

[Rie97b] Riekert, W.-F., Mayer-Föll, R., Wiest, G.; Management of Data and Services in the Environmental Information System (UIS) of Baden-Württemberg; ACM SIGMOD Record, 26 (2), Special Issue on Environmental Information Systems (ed. Günther, O.), 1997.

[SeLe96] Seggelke, J., Lessing, H.; Globales Umweltinformationsnetz: Eckpunkte, Chancen und Gefahren; Informatik für den Umweltschutz, 10. Symposium, Hannover (Hrsg. Lessing, H., Lipeck, U.W.); Metropolis-Verlag, Marburg, 1996.

[SeMo97] Seggelke, J., Mohaupt-Jahr, B.; Der Verweis- und Kommunikationsservice des Umweltbundesamts – Ein Modellfall für das Umwelt-Intranet; Umweltinformatik '97 / Informatique pour l'Environnement '97 (Hrsg. Geiger, W., Jaeschke, A., Rentz, O., Simon, E., Spengler, Th., Zilliox, L., Zundel, T.); Metropolis-Verlag, Marburg, 1997.

[Smi96] Smith, T.; A Brief Update on the Alexandria Digital Library Project; D-LIB The Magazine of Digital Library Research, März 1996; <http://www.dlib.org/dlib/march96/briefings/smith/03smith.html>.

[Toc96] Tochtermann, K.; Organisation und Verwaltung der Verweise im WWW; Internet-basierte Informationssysteme der Bibliotheken, DFG-Workshop an der Universität Bielefeld; <http://www.ub.uni-bielefeld.de/workshop/tochtermann/index.html>, 1996.

[ToRi97] Tochtermann, K., Riekert, W.-F., Wiest, G., Seggelke, J., Mohaupt-Jahr, B.; Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs; Proc. First European Conference on Research and Advanced Technology for Digital Libraries, Pisa, September 1997; Lecture Notes of Computer Science, Springer, 1997.

[USGS 97] <http://www.usgs.gov/>.

[ViLi97] <http://www.w3.org/pub/DataSources/bySubject/Overview.html>.

[Wei95] Weibel S.; Metadata: The Foundations of Resource Description, D-LIB The Magazine of Digital Library Research, Juli 1995; <http://www.dlib.org/dlib/July95/07weibel.html>.

[Z39.50] <http://lcweb.loc.gov/z3950/agency/>.