

Erschließung von Fachinformationen im Internet mit Hilfe von Thesauri und Gazetteers

Wolf-Fritz Riekert^{1,2}, Christian Fuchs² und Gerd Klingler²

Abstract

Die automatisierte Nutzung von Wissensstrukturen wie Thesauri und Gazetteers eröffnet neuartige Methoden zur Erschließung von Fachinformationen im Internet. Leistungsfähige Benutzungsoberflächen ermöglichen eine wirkungsvolle Informationsrecherche über Fach- und Raumbezüge und sind daher insbesondere im Umweltbereich von großem Nutzen. Dies wird an Beispielen verschiedener Entwicklungen demonstriert, die attraktive Alternativen zu der bisher in Metainformationssystemen und Suchmaschinen üblichen Blackbox-Suche aufzeigen.

1. Ausgangspunkt

Die Zunahme von Informationen, die auf dem Internet angeboten werden, hält unvermindert an. Dies trifft in besonderem Maße auch auf den Umweltbereich zu, da gerade jetzt viele Umweltbehörden, Umweltforschungseinrichtungen und Umweltorganisationen ihre Informationen im Internet der Öffentlichkeit zur Verfügung stellen. 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 im Internet 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

¹ Hochschule für Bibliotheks- und Informationswesen, Wolframstr. 34, D-70191 Stuttgart, Email: riekert@hbi-stuttgart.de

² Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW), Postfach 2060, D-89010 Ulm, Email: fuchs | klingler@faw.uni-ulm.de

WWW-Seiten vielfältige Informationsressourcen erreichbar sind, die außer multimedialen Dokumenten auch Daten und Dienstprogramme, z.B. zur Datenbankabfrage umfassen, ermöglichen Suchmaschinen ausgehend von einer einfachen Stichwortangabe 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 ein im WWW eventuell vorhandener Artikel über den Anbau von Apfelsinen auf der Insel Mainau bei Angabe der Stichwörter „Südfrucht“ und „Bodensee“ nicht gefunden werden, da die Suchmaschine über keinerlei Wissen darüber verfügt, daß Apfelsinen Südfrüchte sind und die Insel Mainau im Bodensee liegt.

Die Erschließung 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. Im Umweltbereich hat sich gezeigt, daß es sinnvoll ist, insbesondere drei Arten von Beschreibungselementen für Informationen zu unterscheiden, nämlich *Zeitbezüge*, *Fachbezüge* und *Raumbezüge* (Riekert et al. 1997, Tochtermann et al. 1997). 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.

Die Katalogisierung und Recherche von Informationen hinsichtlich ihrer Fach-, Raum- und Zeitbezüge geschieht in der Regel mit Hilfe von Metainformationssystemen³. Abbildung 1 zeigt mit der Recherchekomponente des Umwelt- und Verkehrs-Informationsservice Baden-Württemberg (UVIS) das Beispiel einer Oberfläche für ein solches Metainformationssystem (Henning et al. 1998). Andere Beispiele finden sich im Prototyp für das Umweltinformationsnetz Deutschland GEIN⁴ (Riekert et al. 1997) und im WWW-UDK, der Web-Version des Umweltdatenkatalogs (Kramer et al. 1996). Die Erschließung von Informationsressourcen erfolgt im Fall von Metainformationssystemen anders als im Fall von Suchmaschinen nicht durch ein Robot-Programm, sondern durch eine Person, in der Regel durch den Systemadministrator oder durch den Anbieter der Originalinformation. Dies hat den Vorteil einer zusätzlichen Qualitätskontrolle, bedeutet aber auch einen erhöhten manuellen Aufwand bei der Aufnahme von Informationsressourcen in den Katalog. Wie im folgenden jedoch noch gezeigt wird, ist es mit Hilfe zusätzlicher Wissensquellen, nämlich mit Hilfe von Thesauri und sogenannter Gazetteers möglich, Suchmaschinen so zu erweitern, daß eine semantische Verarbeitung von Recherchen auch ohne vorangegangene manuelle Katalogisierung weitgehend mög-

³ Der Name Metainformationssysteme rührt daher, daß diese Systeme Informationen über Informationen, sogenannte Metainformationen enthalten.

⁴ Das Akronym GEIN steht für German Environmental Information Network.

lich wird. Darüber hinaus ermöglichen Thesauri und Gazetteers sehr flexible und leistungsfähige Oberflächen zur Katalogisierung und Recherche von Informationsressourcen und können so die Funktionalität von Metainformationssystemen und Suchmaschinen erheblich aufwerten.

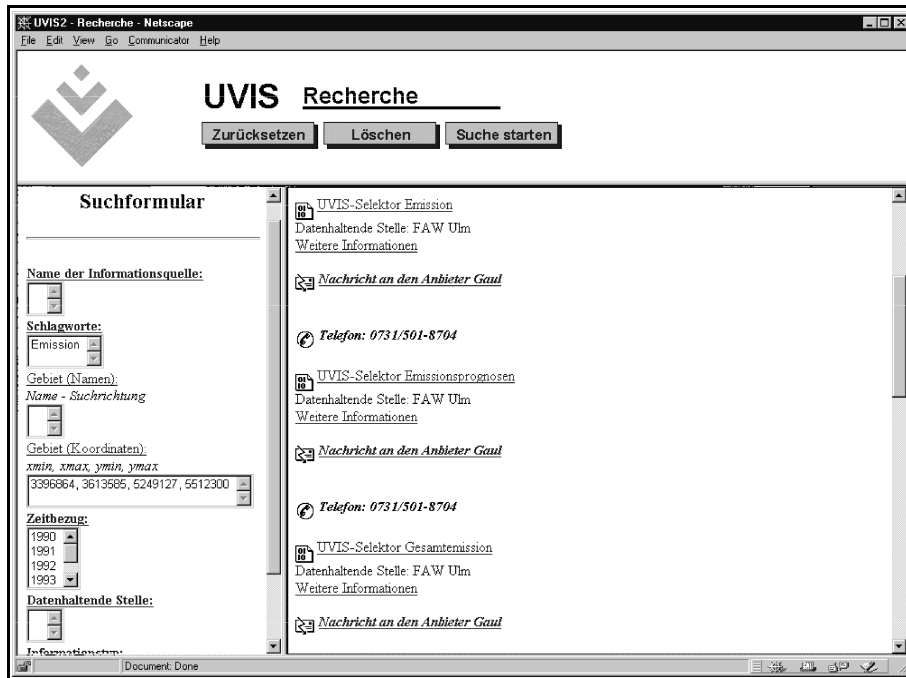


Abbildung 1

Die Recherchekomponente des Umwelt- und Verkehrs-Informationsservice Baden-Württemberg (UVIS) ermöglicht die Suche nach Umweltinformationen hinsichtlich Sachbezug (über Schlagworte), Raumbezug (über Gebietsnamen oder mit der Maus spezifizierte Koordinaten, siehe hierzu auch Abbildung 5) und Zeitbezug. Außerdem können Namen und Typ der Informationsquelle sowie der Namen der datenhaltenden Stelle als Recherchekriterium angegeben werden. Die Ergebnisdarstellung auf der rechten Seite erlaubt es, weitere Informationen abzurufen, auf Originalinformationen durchzugreifen, die im WWW angeboten werden, sowie mit den Informationsanbietern über Email Kontakt aufzunehmen.

2. Verarbeitung von Fachbezügen mit Hilfe eines Thesaurus

Die einfachste und wirkungsvollste Möglichkeit, Fachbezüge zu bezeichnen, geschieht mit Hilfe von Schlagwörtern. Wünschenswert ist hierfür allerdings ein

kontrolliertes Vokabular. Dabei kann insbesondere auf Erfahrungen aus dem Dokumentationsbereich mit sogenannten *Thesauri* zurückgegriffen werden. Ein solcher Thesaurus besitzt zweierlei Funktionen. Zunächst stellt er einen Katalog dar, der die Namen aller Deskriptoren enthält, die für die Verschlagwortung von Informationen genutzt werden können. Darüber hinaus ist ein Thesaurus weitaus mehr als nur ein Schlagwortkatalog dar; denn er behandelt die Deskriptoren 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 Thesaurus festlegen, daß „Südfrucht“ ein Oberbegriff des Begriffs „Orange“ und dieser wiederum ein Synonym des Begriffs „Apfelsine“ ist.

Es ist offenkundig, daß der erste Teil der in Abschnitt 1 angeführten Recherche (nach Informationen mit Fachbezug „Südfrucht“) mit Hilfe des in einem Thesaurus repräsentierten begrifflichen Wissens unterstützt werden kann. Beispiele thesaurusbasierter Systemkomponenten, die dies unterstützen, sind im folgenden beschrieben.

2.1 Unterstützung der Recherche durch einen Thesaurusnavigator

In den meisten existierenden Metainformationssystemen wird derzeit eine sogenannte Blackboxsuche realisiert. Hierbei füllt der Benutzer ein mehr oder weniger umfangreiches Suchformular teilweise oder ganz aus. Das führt schnell zum Ziel, wenn man genau weiß, was man sucht. Andernfalls besteht die Gefahr, daß die Suchkriterien immer weiter eingeschränkt werden, obwohl schon längst keine Treffer mehr erzielt würden.

Deshalb lehnen Benutzer oft Suchformulare ab und fordern einen navigatorischen Zugang über ein thematisch geordnetes hierarchisches Verzeichnis. Ein solches Verzeichnis hat den Vorteil, daß der Benutzer sofort erkennt, zu welchen Themen es Informationen gibt. Zudem ist die Suche über hierarchische Verzeichnisse für die meisten Computerbenutzer ein geläufiger Vorgang.

Da sich ein solches Verzeichnis bei den sich häufig ändernden Metadatenbeständen auf praktikable Weise nicht manuell pflegen läßt, sind hierfür automatische Verfahren gefordert. In einer Arbeit, die am FAW Ulm durchgeführt wurde, wurde der Umweltthesaurus des Umweltbundesamts (Batschi 1994), der in UVIS und im GEIN-Prototyp verwendet wird, für die automatisierte Erzeugung eines solchen Verzeichnisses genutzt.

Bei jeder Änderung des Metainformationsbestands muß das hierarchische Verzeichnis neu generiert werden. Dieses Verzeichnis besteht im wesentlichen aus dem „ausgedünnten“ Thesaurus, der nur noch die Begriffe enthält, die bezüglich des Metainformationsbestands relevant sind. Hierzu wird zunächst ermittelt, welche

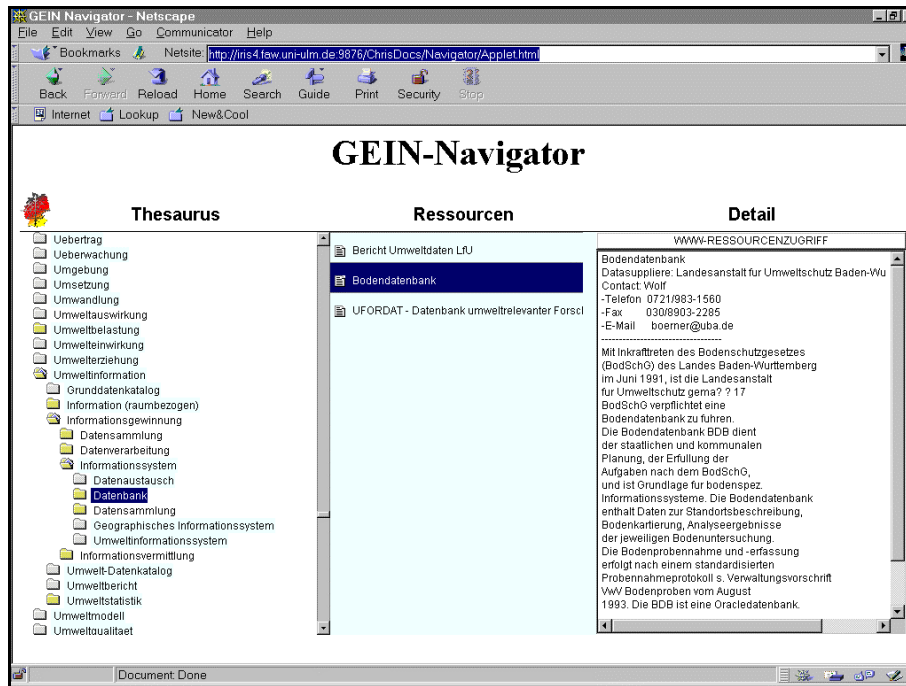


Abbildung 2

Navigatorischer Zugang zu Metainformationen über ein hierarchisches Verzeichnis, das aus den relevanten Begriffen des Umweltthesaurus generiert wurde.

Begriffe aus dem Umweltthesaurus überhaupt für die Katalogisierung von Informationsressourcen genutzt wurden. Zu diesen Begriffen werden rekursiv alle Oberbegriffe bis hin zu den Toplevel-Begriffen (d.h. den Begriffen, die keine Oberbegriffe mehr haben) hinzugenommen und aus der Gesamtheit dieser Begriffe ein polyhierarchisches Verzeichnis erzeugt. Polyhierarchisch bedeutet, daß Begriffe mehrere Oberbegriffe haben können und daher auch als Einträge in diesem Verzeichnis mehrfach auftreten können.

Das aus dem Thesaurus generierte Verzeichnis wird in einem „Navigator“-Fenster dargestellt, das ähnlich wie der Microsoft Windows Explorer aufgebaut ist (Abbildung 2). Die Oberfläche besteht aus drei Teilfenstern. Links werden im Grundzustand alle relevanten Toplevel-Begriffe des Umweltthesaurus aufgelistet. Existieren weitere relevante Unterbegriffe, ist dem Begriff ein gelbes Ordnersymbol zugeordnet. Wird auf das Symbol geklickt, so werden alle Begriffe der darunterliegenden Ebene gezeigt. Begriffen ohne weitere Unterbegriffe ist das gleiche, aber farblose Symbol zugeordnet. Die Namen der Thesaurusbegriffe, die unmittelbar zu mindestens einem Ergebnis führen, sind blau unterlegt. Werden diese durch eina-

ches Anklicken selektiert, so werden im mittleren Teilfenster alle Treffer aufgelistet. Im Beispiel wurden drei Treffer erzielt, wobei das Ergebnis „Bodendatenbanken“ angeklickt wurde, um weitere Metainformationen im rechten Teilfenster abzurufen. Hier erscheinen Informationen wie Kurzbeschreibung, Datenanbieter (Data-supplier), Ansprechpartner mit Telefon, Fax und E-Mail-Adresse, sowie eine ausführlichere Beschreibung der Ressource. Zusätzlich kann durch den Button WWW-Ressourcenzugriff auch direkt zu der hinterlegten Internet-Adresse (URL) gesprungen werden.

Der Navigator wurde in Java implementiert, der Zugriff auf den Thesaurus und die Metadatenbank erfolgt mit Hilfe von JDBC (Java Database Connectivity). Dies macht es möglich, den Navigator aus einem Internetbrowser heraus zu aktivieren, ohne daß auf Nutzerseite zusätzliche Clientsoftware installiert werden muß.

2.2 Thesaurusbasierte Recherche in Suchmaschinen

Wie in den vorangegangenen Abschnitten gezeigt, unterstützen Metainformationssysteme wie auch Suchmaschinen beide die Recherche nach Fachinformationen und den direkten Zugriff auf diese Informationen über Hyperlinks. Unterschiede bestehen jedoch hinsichtlich des Aufwands, der durch die Informationsanbieter bzw. durch die Informationsnutzer jeweils erbracht werden muß.

Metainformationssysteme fordern von den Informationsanbietern einen hohen Arbeitsaufwand für die Katalogisierung (z.B. Verschlagwortung) der Informationsangebote ab. Die Informationsnutzer können jedoch in der Regel in einem gut gepflegten Bestand von Katalogeinträgen mit Hilfe eines problemangemessenen Vokabulars recherchieren, das oft auch alternative Formulierungen (z.B. in Form von Synonymbegriffen oder Oberbegriffen) zuläßt

Suchmaschinen erfordern auf der Anbieterseite keinerlei Aufwand zum Aufbau und zur Pflege des Katalogs, da diese Aufgabe durch das Robot-Programm übernommen wird. Für den Informationsnutzer besteht jedoch, wie in Abschnitt 1 angesprochen, der Nachteil, daß Abfragen ausschließlich als Volltextrecherchen interpretiert werden. Eine inhaltliche Interpretation der Anfragen ist der Suchmaschine nicht möglich, so daß nur solche Informationsressourcen gefunden werden, die den Suchbegriff wortwörtlich enthalten. Andere Formulierungen führen nicht zum Ziel.

Dies war der Anlaß für die Entwicklung der nachfolgend beschriebenen Methode, die teilweise die Vorteile von Metainformationssystemen mit denen von Suchmaschinen verbindet, wobei wiederum ein Thesaurus zum Einsatz kommt. Die Idee besteht darin, der Recherche mit einer Suchmaschine einen Thesaurusmodul vorzuschalten. Der Informationssuchende interagiert mit dem Thesaurusmodul. Ergebnis dieser Interaktion ist die Konstruktion einer Suchanfrage, die vom Thesaurusmodul ggf. noch umformuliert und an die Suchmaschine weitergeleitet wird. Das von der Suchmaschine ermittelte Rechercheergebnis wird unverändert angezeigt.

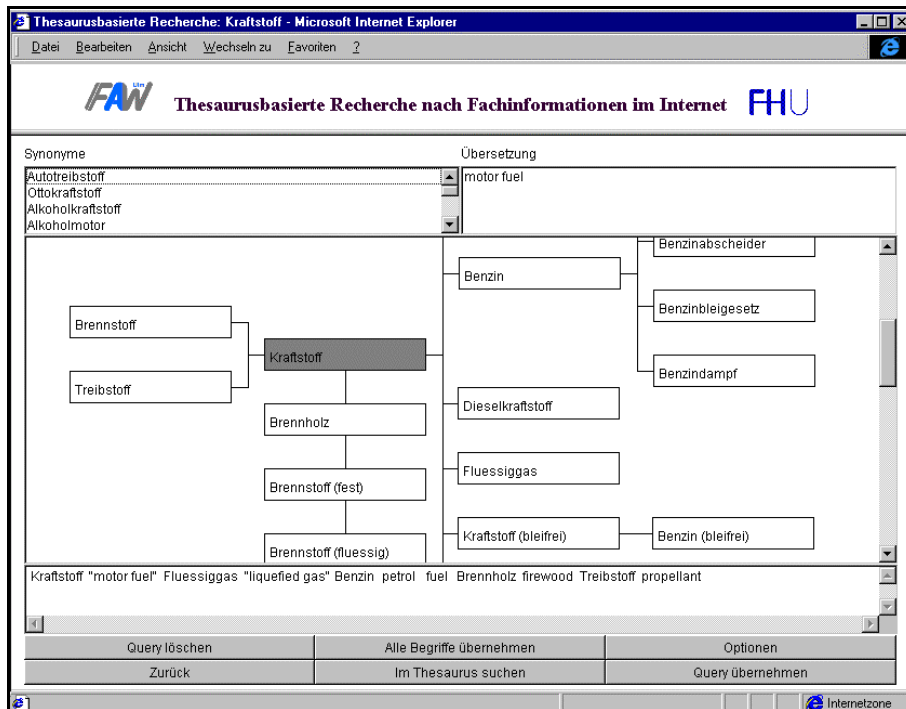


Abbildung 3

Thesaurusbasierte Recherche im Internet: Der Benutzer erkundet den Thesaurus in der Umgebung des Begriffs Kraftstoff. Oberbegriffe im Abstand 1 und Unterbegriffe im Abstand 2 sowie Synonyme und sog. Schwesterbegriffe sind in diese Umgebung mit eingeschlossen, da der Benutzer dies in einem Optionsbogen (Abbildung 4) so spezifiziert hat. Die resultierende Anfrage wird (wie im Optionsbogen angefordert) in fremde Sprachen (z.Zt. nur Englisch) übersetzt und kann über den Button „Suche starten“ auf einer Suchmaschine ausgeführt werden.

Die Vorschaltung des Thesaurusmoduls vor die Suchmaschine ermöglicht zwei neue Arten der Rechercheunterstützung (Abbildung 3):

1. Es besteht die Möglichkeit der Recherche durch Navigation im semantischen Netz, das durch den Thesaurus gegeben ist, entlang der verschiedenen Beziehungen zwischen den Thesaurusbegriffen. Bei dieser Navigation durch den Thesaurus können besuchte Knoten als Suchbegriffe in die Recherche aufgenommen werden und auf diese Weise sukzessive Suchanfragen konstruiert werden.
2. Es besteht die Möglichkeit der Umformulierung von Recherchen. Von den Benutzern angegebene Stichwörter werden in den Thesaurusbegriffen gesucht. Die

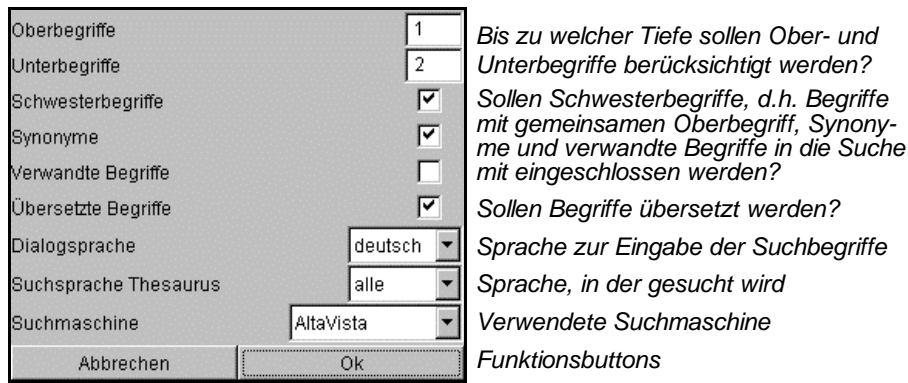


Abbildung 4

Festlegung der Optionen zur thesaurusbasierten Recherche im Internet: In diesem Beispiel werden direkte Oberbegriffe, Unterbegriffe in zwei Stufen, Schwesterbegriffe, Synonyme und übersetzte Begriffe berücksichtigt. Die Spezifikation der Suchbegriffe erfolgt im Dialog auf deutsch, gesucht wird in allen verfügbaren Sprachen mit der Suchmaschine Altavista.

Anfrage wird erweitert, indem die hierbei gefundenen Thesaurusbegriffe um Synonymbegriffe, Unterbegriffe, ggf. auch durch Oberbegriffe, verwandte Begriffe, Schwesterbegriffe usw. ergänzt sowie in fremde Sprachen übersetzt werden, bevor die Anfrage an die Suchmaschine weitergeleitet wird.

Als Thesaurus findet der zweisprachige Umweltthesaurus des Umweltbundesamts Verwendung. Er soll zu gegebener Zeit durch den allgemeinen europäischen multilingualen Umwelthesaurus GEMET ersetzt werden, der nach demselben Datenmodell aufgebaut ist. Der Thesaurusmodul wurde in Form eines Java-Applets entwickelt. Der multilinguale Thesaurus liegt in Form einer relationalen Datenbank entsprechend den Normen ISO 2788 und ISO 5964 vor. Der Zugriff auf diese Datenbank geschieht mit Hilfe von JDBC. Als Suchmaschine wird im Normalfall Altavista (1998) verwendet, jedoch kann das System auch so konfiguriert werden, daß eine andere Suchmaschine benutzt wird. Die Ansteuerung der Suchmaschine erfolgt über das HTTP-Protokoll, das dem WWW zugrunde liegt. Vom Java-Programm aus wird die Suchmaschine genauso angesprochen, als wäre die Anfrage interaktiv in deren Suchformular spezifiziert worden. Eine Nachbearbeitung der Suchergebnisse findet nicht statt; die Ergebnisausgabe wird direkt von der Suchmaschine übernommen.

Die Praxis zeigt, daß Recherchen nach sehr speziellen Fachinformationen im Internet mit bloßen Suchmaschinen oft nur sehr wenige brauchbare Ergebnisse bringen. Durch die mit der beschriebenen Methode mögliche Hinzunahme von weiteren Suchbegriffen aus der thematischen Umgebung des ursprünglichen Be-

griffs und durch die optionale Übersetzung der Begriffe konnte demgegenüber die Ausbeute von Recherchen in vielen Fällen deutlich verbessert werden. Je nach Einstellung der Optionen (Abbildung 4) konnte eine Erhöhung der Anzahl relevanter Ergebnisse um Faktoren zwischen 2 und 100 erzielt werden. Mit dem System steht Informationssuchenden somit ein flexibles, problemangemessenes Recherchewerkzeug zur Verfügung, das den Informationsanbietern keinerlei Arbeitsaufwand auferlegt und daher im besten Maß für den Einsatz im Internet geeignet ist.

3. Verarbeitung von Raumbezügen mit Hilfe eines Gazetteers

Die Interpretation des zweiten Teils der in Abschnitt 1 angeführten Recherche (nach Informationen mit Raumbezug „Bodensee“) setzt eine sogenannte Georeferenzierung der Informationen, das heißt eine Beschreibung der Informationen durch Raumbezüge voraus. Die Spezifikation des Raumbezugs einer Information kann prinzipiell auf zweierlei Weisen erfolgen, nämlich entweder textuell durch Angabe eines geographischen Namens (z.B. im oben genannten Beispiel durch den Namen „Mainau“) oder geometrisch durch die Festlegung von Koordinaten. Letzteres kann durch Eingabe numerischer Werte oder durch Zeigeaktionen mit Hilfe einer Maus erfolgen (Abbildung 5). Prinzipiell sind alle diese Möglichkeiten der Angabe von Raumbezügen sowohl bei der Recherche wie auch bei der Katalogisierung von 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 geographische *Regionen* verstanden werden, ähnlich wie ein Thesaurus als Verwaltungsstruktur für Fachbegriffe aufgefaßt werden kann. Geographische Regionen lassen sich nach unterschiedlichen Typen unterscheiden; wichtige Beispiele sind Verwaltungseinheiten (d.h. Bundesländer, Landkreise, Gemeinden etc.), naturräumliche Gliederungen, Postleitgebiete oder Kartenblattschnitte. Geographische Regionen besitzen einen *Namen* sowie eine *Geometrie* in Form geographischer Koordinaten. Die gleichzeitige Unterstützung beider Arten von Beschreibungselementen ist von besonderer Bedeutung, da es hierdurch möglich ist, frei zwischen geometrischer und textueller Angabe eines Raumbezugs zu wechseln. So kann etwa die geographische Katalogisierung einer Informationsressource textuell erfolgen, die Recherche hingegen geometrisch, wobei der Gazetteer die Übersetzung der Beschreibungselemente vom einen in das andere Bezugssystem ermöglicht

Wie die Begriffe in einem Thesaurus sind auch die Regionen in einem Gazetteer untereinander durch Beziehungen verknüpft. Hierbei sind geometrisch/topologische Beziehungen wie die *Überlappung*, das *Ineinanderenthaltensein* und die gegenseitige *Berührung* von Regionen von besonderer Bedeutung. Diese Beziehungen müs-

sen jedoch nicht explizit gespeichert werden, sondern lassen sich aus der Geometrie der Raumbezüge ableiten.

Geometrisch-topologische Beziehungen ermöglichen eine weitestgehende Freiheit bei der Wahl der Beschreibungselemente für Informationsressourcen. So müssen bei der Recherche nicht dieselben Namen von Regionen zur Angabe des Raumbezugs verwendet werden wie bei der Katalogisierung. Die geometrisch-topologischen Beziehungen zwischen Regionen im Gazetteer können genutzt werden, um zu bestimmen, inwieweit eine Raumbezugsangabe eine andere impliziert.

Die Geometrie einer Region 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 Geometrien, hat jedoch den Vorteil, daß hierfür ein relationales Datenbanksystem ausreicht und kein geographisches Informationssystem (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 über eine Tabelle mit den Regionen assoziieren. Diese Möglichkeit wurde in UVIS und im GEIN-Prototyp gewählt, da in diesen Systemen kein GIS eingesetzt werden sollte. Als dritte Möglichkeit besteht die Verwendung von *Vektorgeometrien*, die eine sehr präzise Repräsentation der Geometrie von Raumbezügen ermöglichen. In diesem Fall empfiehlt es sich, ein echtes GIS zu nutzen.

Der Gazetteer kann nun auf verschiedene Weisen für die Recherche nach Informationen genutzt werden. So ist es mit Hilfe eines *Regionenbrowsers* möglich, Regionen in textueller Form zusammen mit umschließenden oder umschlossenen sowie überlappenden Regionen darzustellen und zu diesen Regionen zu navigieren. Die Namen der aufgesuchten Regionen können bei der Recherche nach Informationen als Suchkriterien übernommen werden. Wie bei der thesaurusbasierten Recherche ist es sinnvoll, über Optionen festzulegen, inwieweit weitere Regionen aufgrund geometrisch/topologischer Beziehungen in die Recherche aufgenommen werden sollen.

Die geometrische Repräsentationsform von Raumbezügen ermöglicht auch *graphische Techniken* der Georeferenzierung. So lassen sich Raumbezüge durch direkte Manipulation mit Hilfe einer Maus spezifizieren (Abbildung 5). Alternativ ist eine Spezifikation von Raumbezügen durch Angabe von Koordinaten möglich.

Auch der Einsatz von Gazetteers ist nicht auf Metainformationssysteme beschränkt. Ähnlich wie für die in Abschnitt 2.2 beschriebene fachbezogene Recherche mit einem Thesaurus können auch 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 Namen von Regionen in einer Recherche um weitere Namen von betroffenen Regionen erweitert und anschließend die Namen all dieser Regionen an die Suchmaschine weitergibt.

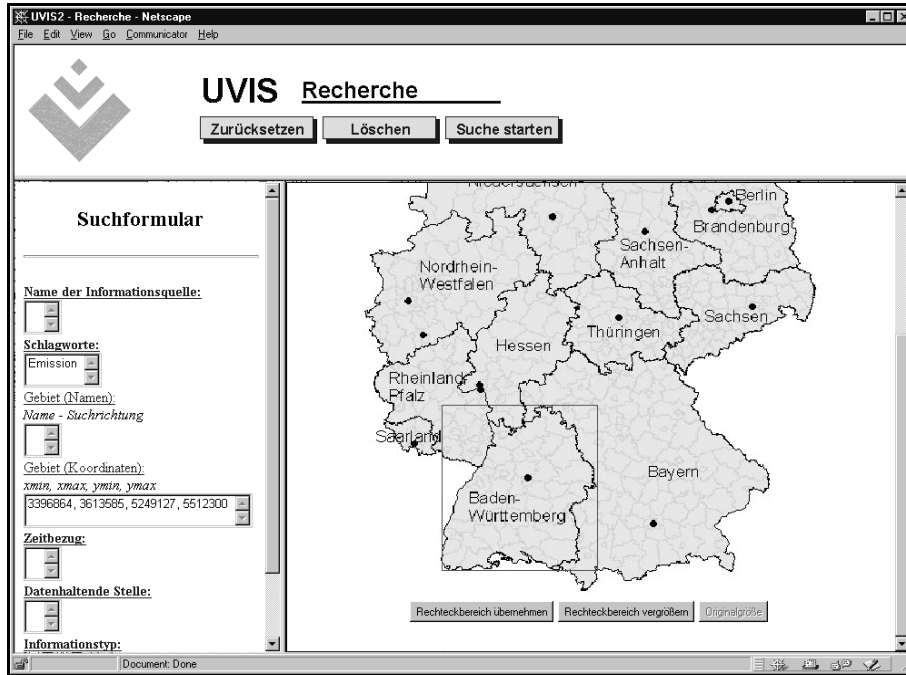


Abbildung 5

Geographische Suche mit der Recherchekomponente des Umwelt- und Verkehrs-Informationsservice Baden-Württemberg (UVIS): Über ein mit der Maus spezifiziertes Suchrechteck werden Regionen aus dem Gazetteer selektiert. Die Recherche in UVIS erbringt dann die Informationsressourcen, die mit den selektierten Regionen in Beziehung stehen.

4. Schlußfolgerungen

Die präsentierten Beispiele zeigen, daß der Einsatz von Wissenstrukturen wie Thesauri und Gazetteers für die Recherche nach Fachinformationen mit Metainformationssystemen und Suchmaschinen große Vorteile erbringt. Insbesondere die verbesserten Möglichkeiten der räumlichen Recherche durch den Einsatz eines Gazetteers sind vor allem im Umweltbereich von großem Nutzen.

Thesauri und Gazetteers ermöglichen die Entwicklung neuer attraktiver Benutzungsoberflächen, die auf der hierarchischen Navigation über den Fachbezug, der automatisierten Umformulierung und Übersetzung von Suchkriterien sowie der raumbezogenen Recherche mittels einer kartographischen Oberfläche beruhen.

Der Pflegeaufwand für diese Lösungen ist dabei vergleichsweise gering. Die Basisinformationen, die einem Thesaurus oder Gazetteer zugrundeliegen, sind im Vergleich zu den Metainformationen in den Katalogen und Suchmaschinen, deren Recherchefunktion sie unterstützen, relativ stabil. Außerdem sind sie unabhängig von konkreten Metainformationsbeständen und können von daher für beliebige Metainformationssysteme und Suchmaschinen genutzt werden. So ist der Umweltthesaurus des Umweltbundesamts bereits heute an vielen Stellen im Einsatz. Ähnliches gilt zunehmend auch für den europäischen multilingualen Umweltthesaurus GEMET.

Deshalb zahlt sich die Arbeit, die in den Aufbau eines Thesaurus oder eines Gazetteer und in die Entwicklung geeigneter Verarbeitungstechniken investiert wird, in vielfachem Maße aus. Entwicklungen, wie sie in diesem Beitrag beschrieben wurden, kommt dementsprechend ein großes Nutzenpotential zu.

Literatur

- Altavista (1998): <http://www.altavista.com>
- Batschi, W.D. (1994): Environmental Thesaurus and Classification of the Umweltbundesamt (German Federal Environmental Agency) Berlin. In: Stancikova, P., Dahlberg, I. (Hrsg.): Environmental Knowledge Organisation and Information Management. INDEKS Verlag, Frankfurt/Main, 1994.
- Henning, I., Wiest, G., Gaul, M., Rainbold, E., Schultze, A., Riekert, W.-F. (1998): UVIS: Umwelt- und Verkehrs-Informationsservice Baden-Württemberg. In: Riekert, W.-F., Tochtermann, K. (Hrsg.): Hypermedia im Umweltschutz. 1. Workshop, Ulm 1998. Metropolis-Verlag, Marburg.
- Kramer, R., Nikolai, R., Keitel, A., Legat, R., Zirm, K. (1996): Enhancing the Environmental Data Catalog UDK for the World Wide Web. In: Lessing, H., Lipeck, U.W. (Hrsg.): Informatik für den Umweltschutz, 10. Symposium, Hannover 1996. Metropolis-Verlag, Marburg.
- Riekert, W.-F., Tochtermann, K., Wiest, G., Gaul, M., Seggelke, J., Mohaupt-Jahr, B. (1997): Fach-, raum- und zeitbezogene Katalogisierung und Recherche von Umweltinformationen auf dem Internet. In: Fuhr, N., Dittrich, G., Tochtermann, K. (Hrsg.): Hypertext – Information Retrieval – Multimedia '97. Proceedings HIM'97. Universitätsverlag Konstanz, 1997.
- Tochtermann, K., Riekert, W.-F., Wiest, G., Seggelke, J., Mohaupt-Jahr, B. (1997): Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs. In: Peters, C., Thanos, C. (eds.): Research and Advanced Technology for Digital Libraries. Proceedings ECDL'97. Springer-Verlag, Berlin.