

Aus:

Informationswirtschaft : Innovation für eine Neue Ökonomie

Hrsg.: Wolf-Fritz Riekert / Martin Michelson,

Wiesbaden : Dt. Univ.-Verl. ; Wiesbaden : Gabler, 2001, S.39-55

## **Techniken zur Erschließung von Informationsressourcen im Internet**

*Wolf-Fritz Riekert*

Die automatisierte Nutzung von Wissensstrukturen wie Thesauri und Gazetteers eröffnet neuartige Methoden zur inhaltlichen Erschließung von Informationsressourcen im Internet. Es lassen sich leistungsfähige Benutzungsoberflächen entwickeln, die eine wirkungsvolle Informationsrecherche über Fach- und Raumbezüge ermöglichen. Dies wird an Beispielen verschiedener Entwicklungen demonstriert, die attraktive Alternativen zu der bisher in Metainformationssystemen und Suchmaschinen üblichen Black-box-Suche aufzeigen.

### **1. Ausgangspunkt**

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 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 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 die Webseite eines Schalldämpferreparaturdienstes in Ottobrunn (Landkreis München) bei Angabe der Stichwörter „Auspuff“ und „Landkreis München“ nicht gefunden werden, da die Suchmaschine über keinerlei Wissen darüber verfügt, dass der Begriff Auspuff in der Alltagssprache anstelle des Fachbegriffs Schalldämpfer verwendet wird und die Gemeinde Ottobrunn im Landkreis München liegt.

Die Erschließung von Informationen über einen Volltextindex reicht also nicht für alle Anwendungsgebiete aus. Es werden deshalb in semantischer Hinsicht tiefer ansetzende Beschreibungselemente für Informationen benötigt. Derartige Beschreibungselemente werden auch als *Metainformationen*<sup>1</sup> bezeichnet, da sie Informationen über Informationen bereithalten. Es hat sich gezeigt, dass es sinnvoll ist, für die inhaltliche Erschließung von Informationsressourcen insbesondere drei Arten von Metainformationen zu unterscheiden, nämlich *Fachbezüge*, *Raumbezüge* und *Zeitbezüge* [Riekert et al. 1997, Tochtermann et al. 1997]. So sieht beispielsweise der weitverbreitete Dublin Core Metadata Standard [Dublin Core Metadata Initiative 2000] neben einem Beschreibungselement für Thema und Stichwörter (DC.SUBJECT) auch ein Beschreibungselement für räumliche und zeitliche Maßangaben (DC.COVERAGE) vor. Die im obigen Beispiel angeführte Recherche zielt auf Informationen mit den Fachbezug „Auspuff“ und Raumbezug „Landkreis München“ ab. Ein Zeitbezug ist in dieser Anfrage nicht spezifiziert; ein solcher könnte beispielsweise durch die zusätzliche Angabe des aktuellen Datums hergestellt werden.

Die Katalogisierung und Recherche von Informationen hinsichtlich ihrer Fach-, Raum- und Zeitbezüge geschieht in der Regel mit Hilfe von sogenannten *Metainformationssystemen*. 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]. 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

---

<sup>1</sup> Es ist auch der Begriff *Metadaten* (englisch metadata) in Gebrauch, der im Folgenden gleichbedeutend mit dem Begriff *Metainformationen* verwendet werden soll.

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 möglich 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.

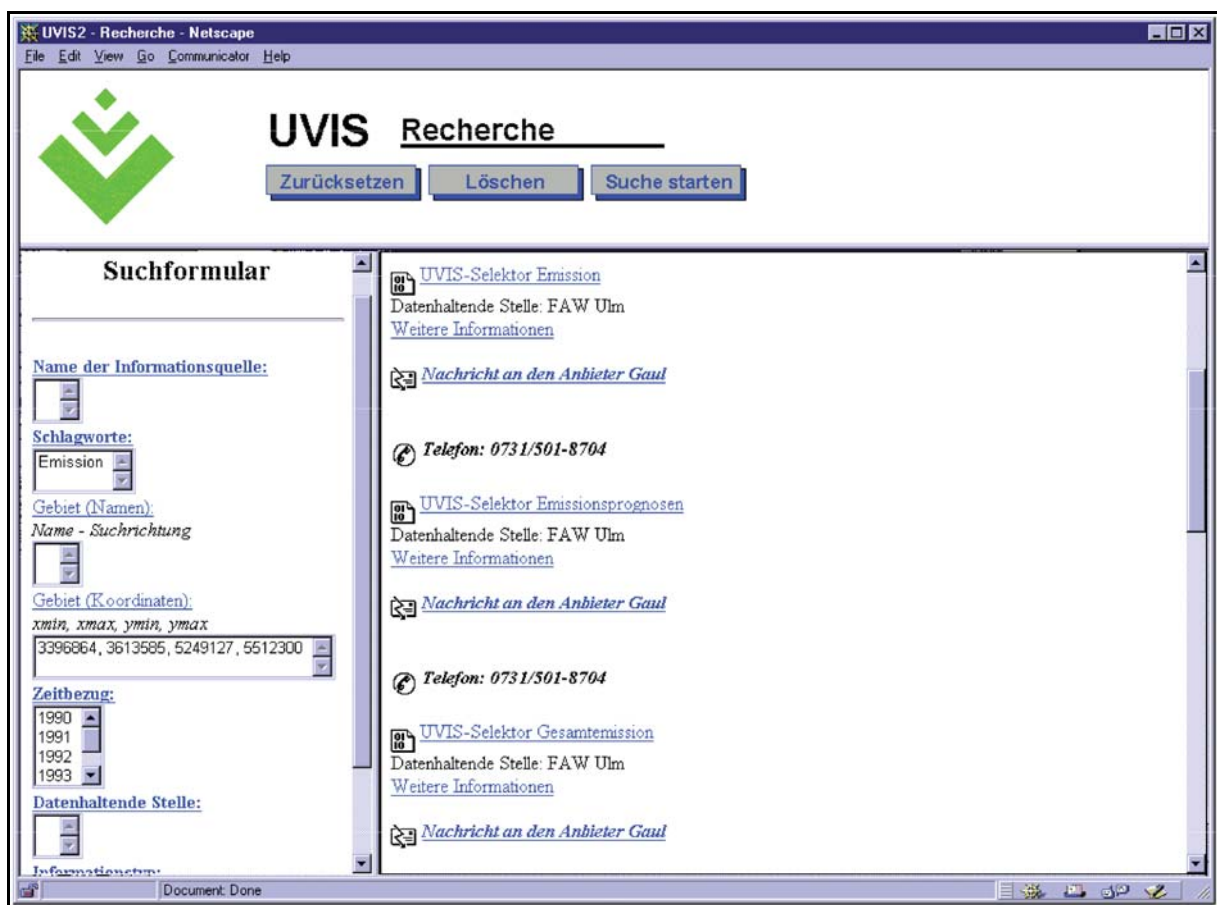


Abbildung 1: Die Recherchekomponente des Umwelt- und Verkehrs-Informationsservice Baden-Württemberg (UVIS) ermöglicht die Suche nach Umweltinformationen hinsichtlich Sachbezug (über Schlagworte), Raumbezug (über Gebietsnamen oder mit der Maus spezifizierte Koordinaten, siehe hierzu auch Abbildung 6 auf Seite 51) 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; 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 „Auspuff“ ein Synonym für „Auspuffanlage“ ist und dieser Begriff wiederum ein Unterbegriff des Begriffs „Schalldämpfer“.

Es ist offenkundig, dass der erste Teil der in Abschnitt 1 angeführten Recherche (nach Informationen mit Fachbezug „Auspuff“) mit Hilfe des in einem Thesaurus repräsentierten begrifflichen Wissens unterstützt werden kann, da dieses die Verbindung zwischen den Deskriptoren, die bei der Verschlagwortung von Informationsressourcen verwendet wurden, und den Begriffen, die ein Informationssuchender verwendet, herstellen kann. Beispiele thesaurusbasierter Systemkomponenten, die dies unterstützen, sind im folgenden beschrieben.

### 2.1 Aufbereitung eines Thesaurus als Hypertext

In einem frühen Versuch aus dem Jahr 1994 wurde der polyhierarchische Umweltthesaurus des Umweltbundesamts [Batschi 1994] als WWW-Hypertext aufbereitet und zur Erschließung eines ebenfalls als Hypertext vorliegenden Berichts genutzt. Abbildung 2 gibt ein Beispiel für die Nutzung des Thesaurus bei der Recherche nach Informationen im Bericht „Umweltdaten“ der Landesanstalt für Umweltschutz Baden-Württemberg [1992]. Die Verschlagwortung der Berichtsseiten mit Deskriptoren aus dem Thesaurus erfolgte automatisch mit Hilfe eines einfachen, aber effizienten Textvergleichsprogrammes [Gaul 1995, Riekert 1996].

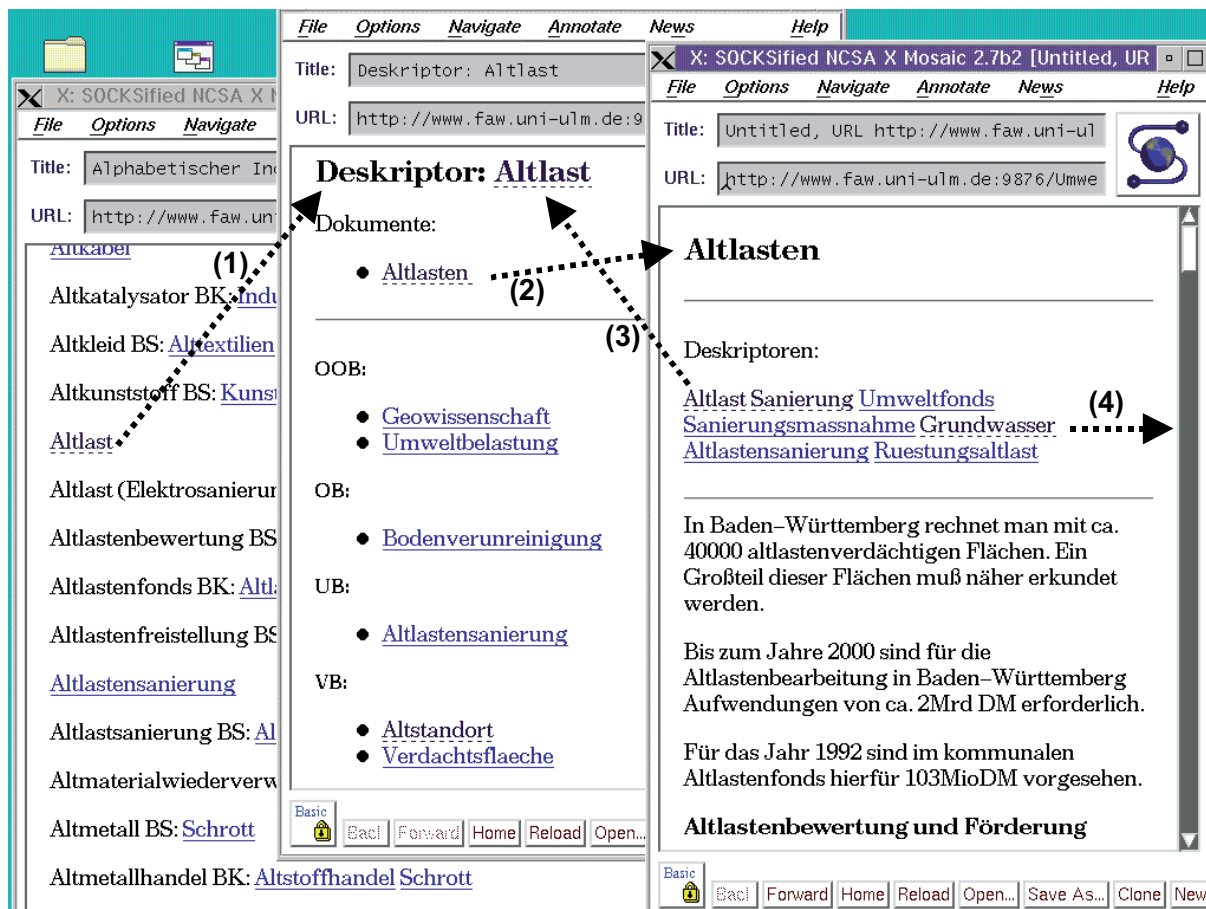


Abbildung 2: Thesaurusbasierter Zugriff zu einem als WWW-Hypertext vorliegenden Bericht. Im alphabetischen Index (linkes Fenster) sind die Namen von Deskriptoren (Thesaurusbegriffen) sowohl direkt wie auch als Synonyme (BS) und Synonymkombinationen (BK) aufgeführt. Durch Anklicken eines Eintrags (hier: *Altlast*) gelangt man zu einem Hypertextdokument, das den entsprechenden Deskriptor zeigt (1). Von dort aus ist es möglich, entweder zu anderen Deskriptoren – Oberbegriffen (OB, OOB), Unterbegriffen (UB) und verwandten Begriffen (VB) – zu navigieren oder zu einem Berichtsabschnitt (hier: *Altlasten*), der mit dem betreffenden Deskriptor verschlagwortet ist (2). Die Deskriptoren, mit denen ein solcher Berichtsabschnitt verschlagwortet ist, sind nach dessen Titel in Form von Hyperlinks aufgeführt. So ist es beispielsweise möglich, wieder zum Deskriptor *Altlast* zurückzukehren (3) oder auch zu einem anderen Schlagwort, beispielsweise *Grundwasser*, überzugehen (4).

## 2.2 Unterstützung der Recherche durch einen Thesaurusnavigator

Der im vorigen Abschnitt beschriebene Ansatz besteht darin, einen statischen Hypertext aufzubauen, der den Thesaurus und die durch den Thesaurus inhaltlich erschlossenen Dokumentseiten umfasst. Der Vorteil dieses Vorgehens besteht darin, dass den Benutzern über die Hyperlinks ein navigatorischer Zugang zu den Informationen er-

möglichst wird. Dieser Informationszugang ist intuitiv sehr leicht zu erfassen und dadurch recht benutzerfreundlich.

Da es sich jedoch um eine statische Lösung handelt, muss der Hypertext stets neu generiert werden, wenn immer sich die Inhalte ändern. Vor allem für die großen verteilten Dokumentstrukturen, die durch das Web gebildet werden, ist diese Technik nicht auf sinnvolle Weise nutzbar. Deshalb geht man in der Regel zu den eingangs erwähnten Metainformationssystemen über, die einen Thesaurus zusammen mit den Metainformationen der interessierenden Informationen in einer Datenbank abspeichern. Die im Metainformationssystem katalogisierten Informationsressourcen existieren in verteilter Form außerhalb des Metainformationssystems.

In den meisten existierenden Metainformationssystemen wird ähnlich wie in dem in Abbildung 1 dargestellten System UVIS eine sogenannte Blackboxsuche realisiert. Hierbei füllt der Benutzer ein mehr oder weniger umfangreiches Suchformular teilweise oder ganz aus. Das kann schnell zum Ziel führen, wenn man genau weiß, was man sucht. Wenn jedoch ein solches System nur gelegentlich benutzt wird, 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 auch in Metainformationssystemen 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 unter der Leitung des Autors am FAW Ulm durchgeführt wurde, wurde der Umweltthesaurus des Umweltbundesamts, der auch in der vor genannten statischen Lösung sowie im System UVIS 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 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 Verzeich-

nis erzeugt. Polyhierarchisch bedeutet, dass Begriffe mehrere Oberbegriffe haben können und daher auch als Einträge in diesem Verzeichnis mehrfach auftreten können.

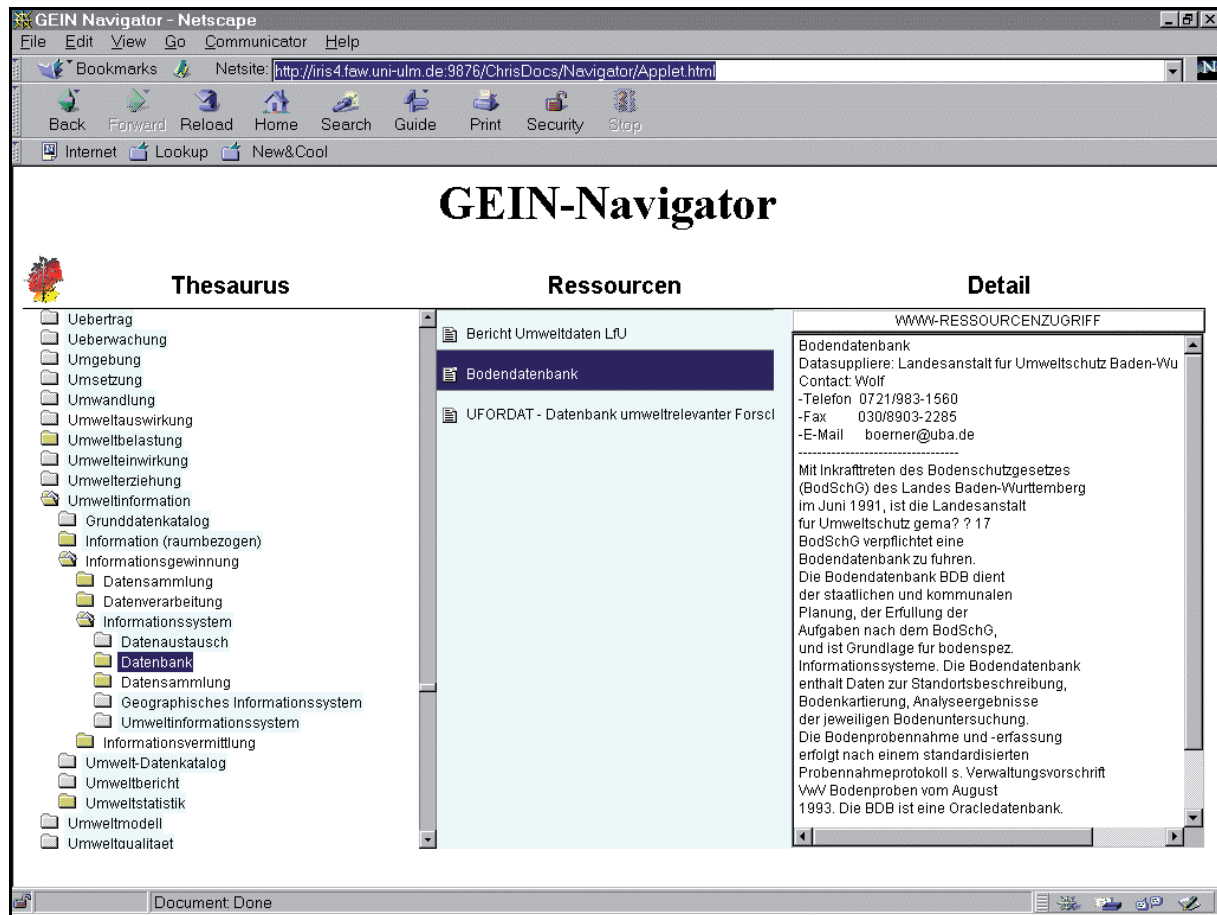


Abbildung 3: Navigatorischer Zugang zu Metainformationen über ein hierarchisches Verzeichnis, das aus den relevanten Begriffen des Umweltthesaurus generiert wurde.

Das aus dem Thesaurus generierte Verzeichnis wird in einem „Navigator“-Fenster dargestellt, das ähnlich wie der Microsoft Windows Explorer aufgebaut ist (Abbildung 3). 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 darunter liegenden 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 einfaches Anklicken selektiert, so werden im mittleren Teilfenster alle Treffer aufgelistet. Im Beispiel wurden drei Treffer erzielt, wobei das Ergebnis „Bodendatenbanken“ angeklickt wurde, um weitere Metain-

formationen im rechten Teilfenster abzurufen. Hier erscheinen Informationen wie Kurzbeschreibung, Datenanbieter, 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.

### 2.3 Thesaurusbasierte Recherche in Suchmaschinen

Wie in den vorangegangenen Abschnitten gezeigt, unterstützen Metainformationssysteme wie auch Suchmaschinen beide die Recherche nach Informationsressourcen 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.



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 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 fand auch hier wieder der Umweltthesaurus des Umweltbundesamts Verwendung, der die beiden Sprachen Deutsch und Englisch unterstützt. Alternativ kann der allgemeine europäische multilinguale Umweltthesaurus GEMET verwendet werden, der nach demselben Datenmodell aufgebaut ist, aber ein etwas geringeres Vokabular bereitstellt. Dieses Datenmodell wurde in einer relationalen Datenbank entsprechend den Normen ISO 2788 und ISO 5964 realisiert. Der Thesaurusmodul wurde in Form eines Java-Applets entwickelt. Der Zugriff auf die Datenbank geschieht mit Hilfe der Schnittstellensoftware Java Database Connectivity (JDBC).

Als Suchmaschine wird Altavista [2000] 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 Informationsressourcen im Internet mit bloßen Suchmaschinen oft nur sehr wenige brauchbare Ergebnisse erbringen. Durch die mit der beschriebenen Methode mögliche Hinzunahme von weiteren Suchbegriffen aus der thematischen Umgebung des ursprünglichen Begriffs 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 5) konnte eine Erhöhung der Anzahl relevanter Ergebnisse um Faktoren zwischen 2 und 100 erzielt werden. Mit dem System steht Informationssuchenden so-

mit ein flexibles, problemangemessenes Recherchewerkzeug zur Verfügung, das den Informationsanbietern keinerlei Arbeitsaufwand auferlegt und daher im besten Maß für den Einsatz im Internet geeignet ist.

The screenshot shows a web browser window titled "Thesaurusbasierte Recherche: Kraftstoff - Microsoft Internet Explorer". The page header includes the logos for "FAW" and "FHU" and the text "Thesaurusbasierte Recherche nach Fachinformationen im Internet".

The main content area is divided into two sections: "Synonyme" and "Übersetzung".

**Synonyme:** A list of terms including "Autotreibstoff", "Ottokraftstoff", "Alkoholkraftstoff", and "Alkoholmotor".

**Übersetzung:** The term "motor fuel" is displayed.

**Thesaurus Diagram:** A hierarchical tree structure centered on "Kraftstoff".

- Level 1 (Oberbegriffe):** "Brennstoff" and "Treibstoff" are connected to "Kraftstoff".
- Level 2 (Unterbegriffe):** "Brennstoff" branches into "Brennholz" and "Brennstoff (fest)". "Treibstoff" branches into "Brennstoff (fest)" and "Brennstoff (fluessig)".
- Level 3 (Schwesterbegriffe):** "Benzin", "Dieselkraftstoff", "Fluessiggas", and "Kraftstoff (bleifrei)" are connected to "Kraftstoff".
- Level 4 (Schwesterbegriffe):** "Benzinabscheider", "Benzinbleigesetz", and "Benzindampf" are connected to "Benzin". "Benzin (bleifrei)" is connected to "Kraftstoff (bleifrei)".

**Translation:** A text box at the bottom of the diagram area contains the translation: "Kraftstoff "motor fuel" Fluessiggas "liquefied gas" Benzin petrol fuel Brennholz firewood Treibstoff propellant".

**Navigation Buttons:** A row of buttons at the bottom includes "Query löschen", "Alle Begriffe übernehmen", "Optionen", "Zurück", "Im Thesaurus suchen", and "Query übernehmen".

Abbildung 4: 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 5) 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.

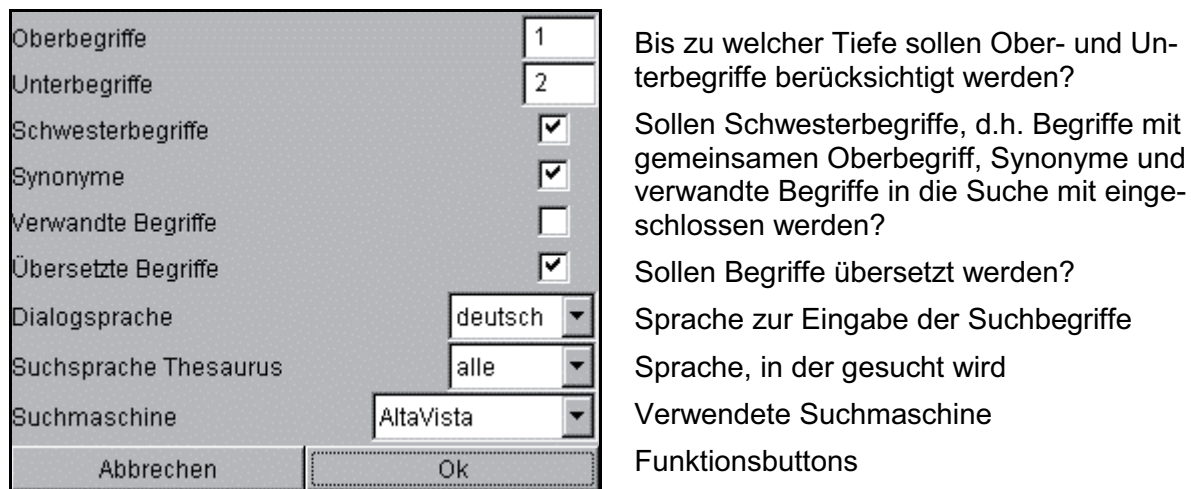


Abbildung 5: Festlegung der Optionen zur thesaurusbasierten Recherche im Internet: In diesem Beispiel werden direkte Oberbegriffe, Unterbegriffe in zwei Stufen, Schwesterbegriffe, Synonyme und übersetzte Begriffe berücksichtigt. Die Spezifikation der Suchbegriffe erfolgt im Dialog auf deutsch, gesucht wird in allen verfügbaren Sprachen mit der Suchmaschine Altavista.

### 3. Verarbeitung von Raumbezügen mit Hilfe eines Gazetteers

Die Interpretation des zweiten Teils der in Abschnitt 1 angeführten Recherche (nach Informationen mit Raumbezug „Landkreis München“) 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 die Bezeichnung „Landkreis München“) oder geometrisch durch die Festlegung von Koordinaten. Letzteres kann durch Eingabe numerischer Werte oder durch Zeigeaktionen mit Hilfe einer Maus erfolgen (Abbildung 6). 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 Abstimmbezirke. 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 müssen 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. Im eingangs angeführten Beispiel bedeutet dies, dass bei Angabe des geographischen Namens „Landkreis München“ auch Informationsressourcen gefunden werden können, die mit der Stadt Ottonbrunn assoziiert sind, da die eine Region die andere enthält.

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 im bereits erwähnten System UVIS gewählt, da in diesem System 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 6). Alternativ ist eine Spezifikation von Raumbezügen durch Angabe von Koordinaten möglich.

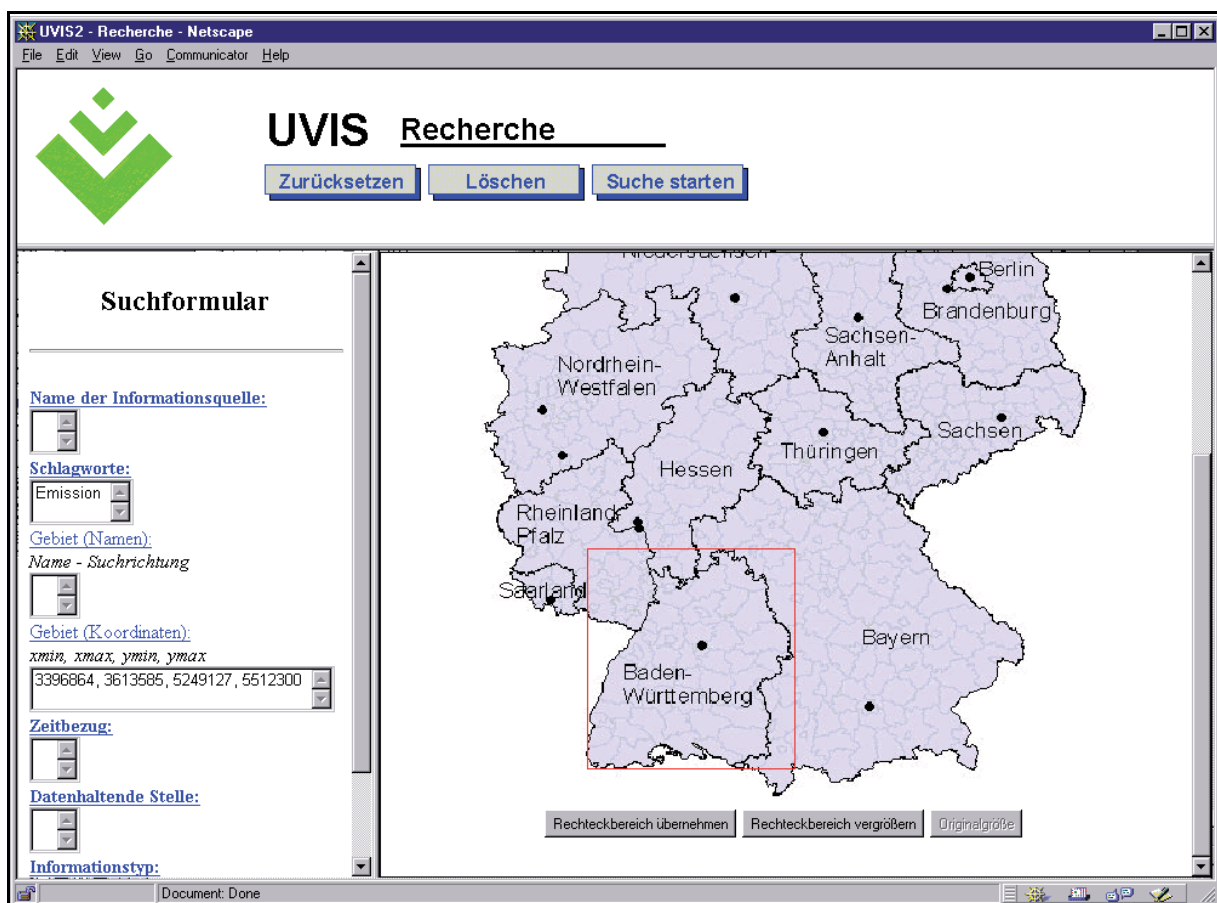


Abbildung 6: Geographische Suche mit der Recherchekomponente des Umwelt- und Verkehrs-Informationsservice Baden-Württemberg (UVIS): Über ein mit der Maus spezifiziertes Suchrechteck werden Regionen aus dem Gazetteer selektiert. Die Recherche in UVIS erbringt dann die Informationsressourcen, die mit den selektierten Regionen in Beziehung stehen.

Auch der Einsatz von Gazetteers ist nicht auf Metainformationssysteme beschränkt. Ähnlich wie für die in Abschnitt 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.

#### **4. Location-based Services in mobilen Netzen**

Das Internet als ein globales Verbundnetz ist derzeit dabei, seinen Siegeszug im drahtlosen Bereich fortzusetzen und die modernen Mobilfunknetze als Teilnetze zu integrieren. Dies ist kein Wunder; denn der Anwendungsbereich der modernen Mobilfunknetze geht weit über die bloße Telefonfunktion hinaus. Bereits der heutige Mobilfunkstandard GSM (Global System for Mobile Communications) erlaubt eine Datenkommunikation mit einer Übertragungsrate von 9600 Bit pro Sekunde. Dadurch lassen sich Notebook- und Palmtop-Computer schon seit geraumer Zeit an das Internet an-koppeln. Mobiltelefone, die für das Übertragungsprotokoll WAP (Wireless Application Protocol) ausgerüstet sind, stellen bereits heute eine Art abgespeckter Internet-Browser-Funktionalität bereit. Die elektronischen Marktplätze beginnen sich in Form eines sogenannten Mobile Commerce auf die drahtlosen Netze auszuweiten.

Die im Sommer 2000 für fast 100 Milliarden DM versteigerten deutschen Lizenzen für das Mobilfunksystem UMTS (Universal Mobile Telecommunication System) werden künftig Datenraten bis zu 2 Millionen Bit pro Sekunde ermöglichen; das ist 32-mal so viel wie in der modernen ISDN-Telefonie üblich. Der Return of Investment kann nur durch Einnahmen für hochwertige breitbandige Netzwerkdienste erbracht werden. Die wirtschaftlichen Erträge durch Sprachtelefonie im herkömmlichen Sinn werden hingegen deutlich abnehmen; hierfür wird bereits der Nulltarif vorhergesagt. Mobiltelefone werden sich in mobile multimediale Terminals wandeln, die alle Arten von Informations- und Kommunikationsbedürfnissen decken, über die Käufe und Geldgeschäfte abgewickelt werden, und die nicht zuletzt als Unterhaltungsmedium dienen [Wienand 2000].

Eine ganz besondere Rolle werden in diesem Kontext sogenannte „Location-based Services“ spielen, das sind Netzwerkdienste, welche die aktuelle geographische Position des Kunden berücksichtigen. Durch Auswertung der Laufzeiten von Mobilfunksignalen ist es ähnlich wie mit dem Global Positioning System (GPS), das den Navi-

gationssystemen gutausgestatteter Limousinen zugrunde liegt, auf zum Teil äußerst exakte Weise möglich, den aktuellen Standort eines Mobilfunkteilnehmers zu ermitteln. Diese Positionsangabe lässt sich für eine geographische Suche nach Informationsressourcen nutzen. Die Suche selbst basiert dann auf den im vorigen Abschnitt beschriebenen Techniken.

So sind bereits heute Dienste für WAP-fähige Mobiltelefone in der Entwicklung, die es erlauben, mit wenigen Tastendrücken standortsbezogene Informationen abzurufen wie z.B. die Wegbeschreibung zum nächstgelegenen Bahnhof und die nächste Zugverbindung von dort nach Hause [Balsiger 2000]. Notrufe können automatisch an die nächstgelegene Polizeiwache vermittelt werden. Auch die oben gestellte Frage nach dem nächstgelegenen Auspuffdienst ließe sich mit dieser Technik auf einfache Weise beantworten. Mit den künftigen breitbandigen Mobilfunkstandards werden dabei auch hochauflösende kartographische Darstellungen der Abfrageergebnisse möglich werden.

Mobilfunkanbieter erwarten ein rapides Wachstum von Location-based Services in den nächsten fünf Jahren. Wienand [2000] zitiert eine Studie der ARC Group mit dem Titel Wireless Internet aus dem Jahre 1999, die für das Jahr 2004 mehr als 400 Millionen Nutzer solcher Services vorhersagt.

## 5. Schlussfolgerungen

Die präsentierten Beispiele zeigen, dass der Einsatz von Wissensstrukturen wie Thesauri und Gazetteers für die Recherche nach Informationsressourcen mit Metainformationssystemen und Suchmaschinen große Vorteile erbringt. Insbesondere die verbesserten Möglichkeiten der räumlichen Recherche durch den Einsatz eines Gazetteers sind für das zukunftssträchtige Gebiet der Location-based Services 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 zugrunde liegen, 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 Um-

weltbundesamts bereits heute an vielen Stellen im Einsatz. Ähnliches gilt zunehmend auch für den Gazetteer des Umweltbundesamts, den sogenannten Geothesaurus [Riekert/Treffler 2000].

Deshalb zahlt sich die Arbeit, die in den Aufbau eines Thesaurus oder eines Gazetteers 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, insbesondere im Hinblick auf die sich bildenden elektronischen und mobilen Märkte.

## Literatur

- Altavista [2000]: <http://www.altavista.com>
- Balsiger, P. [2000]: Mogid – „Mobile Geo-depended Information Demand“. 3. Seminar GIS im Internet/Intranet. Seminarunterlagen. Universität der Bundeswehr München.
- 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.
- Dublin Core Metadata Initiative [2000]: <http://purl.oclc.org/dc/>.
- Gaul, M. [1995]: Thesaurusgestützter Zugriff zu Umweltberichten in einem netzübergreifenden Hypertextsystem. FAW-Forschungsbericht FAW-TR-95006. FAW Ulm.
- 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.
- Landesanstalt für Umweltschutz Baden-Württemberg [1992]: Umweltdaten 91/92, Karlsruhe.
- Riekert, W.-F. [1996]: Thesaurusgestützter Zugriff zu Umweltinformationen in einem netzübergreifenden Hypermediasystem, in: Güttler, R. / Geiger, W.: Integration von Umweltdaten, 3. Workshop, Metropolis-Verlag, Marburg.
- Riekert, W.-F., Fuchs, Ch., Klingler, G. [1999]: Erschließung von Fachinformationen im Internet mit Hilfe von Thesauri und Gazetteers. In: Dade, Ch., Schulz, B. (Hrsg.): Management von Umweltinformationen in vernetzten Umgebungen. 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.
- Riekert, W.-F., Treffler, P. [2000]: Georeferenzierung als Mittel zur Erschließung von Fachinformationen in Internet und Intranet. In: Umweltinformatik 2000, Tagungsband. Metropolis-Verlag, Marburg.
- 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.

Wienand, P. [2000]: WAP und HSCSD, GPRS, EDGE – Das Internet wird mobil. 3. Seminar GIS im Internet/Intranet. Seminarunterlagen. Universität der Bundeswehr München.