

# **Thesaurusgestützter Zugriff zu Umweltinformationen in einem netzübergreifenden Hypermediasystem**

Wolf-Fritz Riekert<sup>1</sup>

## **Zusammenfassung**

Im Verbundprojekt GLOBUS<sup>2</sup>, einem Forschungsvorhaben im Rahmen des Umweltinformationssystems (UIS) Baden-Württemberg, arbeitet das Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW) an der Entwicklung von Techniken, die den Zugriff zu Umweltinformationen im netzübergreifenden Hypermediasystem World-Wide Web (WWW) erleichtern. Diese Arbeit steht im engen Zusammenhang mit dem Forschungsprojekt INTEGRAL<sup>3</sup>, in dem das FAW Ulm heterogene, verteilte Komponenten des UIS Baden-Württemberg integriert. WWW erlaubt hierbei die rechnerübergreifende Navigation hin zu den gesuchten Informationen, die dem Benutzer als Hypermediadokumente zur Verfügung gestellt werden. Dabei hat sich gezeigt, daß die Navigation über Hyperlinks alleine nicht ausreicht, um einen komfortablen Zugang zu den Diensten und Daten des UIS zu ermöglichen.

Im Rahmen der hier beschriebenen Arbeiten werden deshalb Retrievaltechniken entwickelt, die es erlauben, mit Hilfe eines Thesaurus und weiterer Metainformationen direkt an die gesuchten Informationen zu gelangen. Eine erste Anwendung dieser Techniken bestand in der Erschließung des Berichts Umweltdaten 91/92 des Umweltministeriums und der Landesanstalt für Umweltschutz (LfU) Baden-Württemberg mit Hilfe des Umweltthesaurus des Umweltbundesamts. In der gegenwärtigen Phase der Arbeiten wird die Recherchefunktionalität erweitert und auf den Zugriff zu allgemeinen verteilten Diensten und Daten im UIS ausgedehnt.

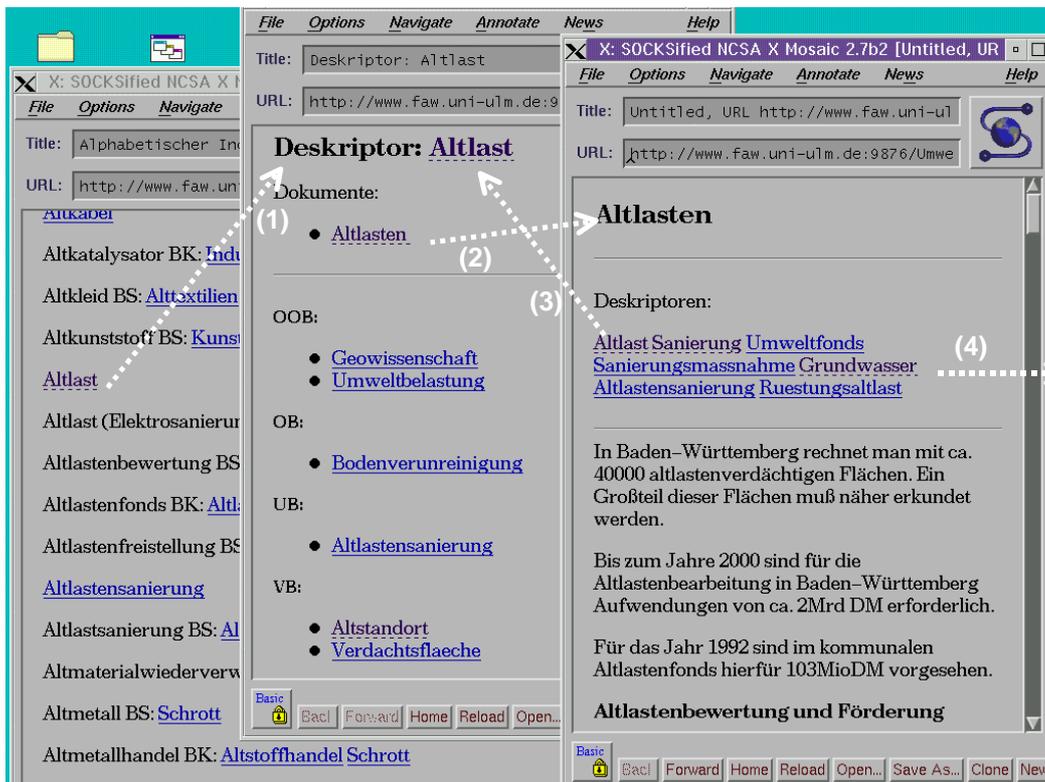
---

<sup>1</sup> Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW) an der Universität Ulm, Postfach 2060, D-89010 Ulm

Tel.: (0731) 501-500, Fax: (0731) 501-999, mailto: riekert@faw.uni-ulm.de

<sup>2</sup> Das Projekt GLOBUS (Konzeption und prototypische Realisierung einer aktiven Auskunftskomponente für globale Umwelt-Sachdaten) wurde vom Umweltministeriums Baden-Württemberg in Auftrag gegeben. Projektpartner waren das Institut für Photogrammetrie und Fernerkundung (IPF) der Universität Karlsruhe, das Institut für Kernenergetik und Energiesysteme (IKE) der Universität Stuttgart sowie das Forschungszentrum Informatik (FZI) an der Universität Karlsruhe. In der Projektphase II (ab Anfang 1995) kam noch das Forschungszentrum Karlsruhe (FZK) hinzu.

<sup>3</sup> Das Projekt INTEGRAL (Integration heterogener Einzelkomponenten des UIS Baden-Württemberg) wurde vom Umweltministeriums Baden-Württemberg in Auftrag gegeben. Das IKE, Universität Stuttgart ist im Rahmen eines Unterauftrags beteiligt.



**Abbildung 1:** Thesaurusbasierter Zugriff zu einem Umweltbericht. Im alphabetischen Index (linkes Fenster) sind die Namen von Thesaurusbegriffen sowohl direkt als auch unter ihren Synonymen (BS) und Synonymkombinationen (BK) aufgeführt. Durch Anklicken eines Eintrags (hier: *Altlast*) gelangt zu einem Dokument, das den entsprechenden Thesaurusbegriff (hier als „Deskriptor“ bezeichnet) zeigt (1). Von dort aus ist es möglich, entweder zu Oberbegriffen (OB, OOB), Unterbegriffen (UB) und verwandten Begriffen (VB) zu navigieren oder zu einem Berichtsabschnitt (hier: *Altlasten*), der mit dem betreffenden Begriff verschlagwortet ist (2). Die Begriffe, mit denen ein solcher Berichtsabschnitt verschlagwortet ist, sind nach dessen Titel in Form von Hyperlinks aufgeführt. So ist es beispielsweise möglich, wieder zum Begriff *Altlast* zurückzukehren (3) oder auch zu einem anderen Schlagwort, beispielsweise *Grundwasser*, überzugehen (4). Auf diese Weise kann ein verschlagworteter Berichtsabschnitt eine semantische Beziehung zwischen Thesaurusbegriffen (hier: *Altlast* und *Grundwasser*) herstellen, die nicht notwendigerweise im Thesaurus selbst reflektiert ist.

## Thesaurusbasierter Zugriff zu Umweltberichten

Erstes Anwendungsbeispiel der vom FAW entwickelten thesaurusbasierten Zugriffstechniken war die Erschließung des Berichts Umweltdaten 91/92 des Umweltministeriums und der Landesanstalt für Umweltschutz (LfU) Baden-Württemberg. Dieser Umweltbericht wurde von unserem Projektpartner, dem Institut für Photogrammetrie und Fernerkundung der Universität Karlsruhe, für das WWW aufbereitet.

Im FAW wurde dann der Thesaurus des Umweltbundesamts genutzt, um diesen Hypertext zu erschließen. Die Begriffe aus diesem Thesaurus wurden dabei zum einen für die Verschlagwortung des Umweltberichts genutzt. Zum andern wurde der Thesaurus selbst als Hypertext aufbereitet.<sup>4</sup>

Für den Informationssuchenden bestehen insgesamt drei Möglichkeiten des Zugangs zu den Dokumentseiten des Umweltberichts. Die erste Möglichkeit besteht im hierarchischen Zugang zu den Informationen über ein systematisches Kapitelverzeichnis und zugehörige Unterverzeichnisse. Hierdurch ist der Benutzer allerdings auf die Systematik des Berichtserstellers eingeschränkt. Die zweite Möglichkeit besteht über ein Volltextsuchprogramm, das über eine WWW-Formularseite gestartet werden kann. Vorteil dieses Verfahrens ist die Möglichkeit beliebige Suchmuster verwenden zu können, Nachteil ist die fehlende Berücksichtigung von Synonymbegriffen.

Die dritte Möglichkeit besteht in der Suche über den Thesaurus. Über einen alphabetischen Index, der als Hypertext präsentiert wird, ermittelt der Benutzer einen interessierenden Begriff und läßt sich diesen als Hypertextdokument zeigen. In diesem Hypertextdokument sind die Titel aller mit dem betreffenden Begriff oder mit seinen Synonymen verschlagworteten Berichtsseiten aufgeführt. Durch Anklicken der Titel können die Berichtsseiten unmittelbar erreicht werden. Sollten die mit dem Begriff assoziierten Berichtsseiten nicht interessieren, so besteht die Möglichkeit zu anderen Begriffen (Oberbegriffen, Unterbegriffen, verwandten Begriffen) zu navigieren und die dort vermerkten Berichtsseiten zu inspizieren. Darüber hinaus ist es auch möglich, von einem gefundenen Berichtsdocument über Hyperlinks zu den Hypertextseiten der Begriffe zu gelangen, mit denen sie verschlagwortet sind, so daß eine freie Navigation zwischen Begriffen und Berichtsseiten unterstützt wird (Abbildung 1).

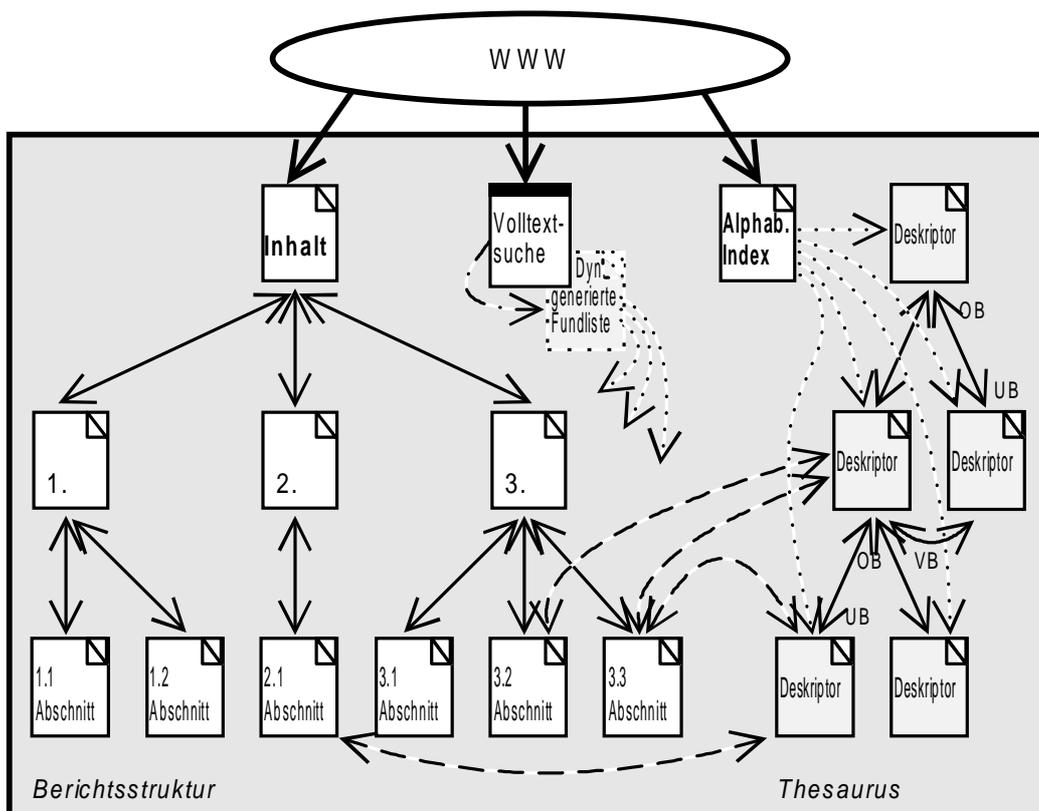
## **Realisierung des thesaurusgestützten Zugriffs**

### **Aufbereitung von Berichten als Hypertext**

Der Bericht Umweltdaten 91/92 (Umweltministerium Baden-Württemberg 1991), der als erstes Anwendungsbeispiel für die beschriebene Entwicklung diente, lag bereits in digitaler Form vor, und zwar im Format des Desktop Publishing Systems Ventura Publisher. Das Institut für Photogrammetrie und Fernerkundung (IPF) der Universität Karlsruhe übernahm die Aufgabe, den derart vorliegenden Text in einen hierarchischen Hypertext zu konvertieren. Mit Hilfe eines eigens für diesen Anwendungsfall entwickelten Verfahrens wurde der Bericht in einzelne Hypertext-Seiten im HTML-Format zerlegt, das dem WWW zugrunde liegt (Abbildung 2). Die einzelnen Unterkapitel des Berichts bilden dabei eigene Hypertext-Seiten,

---

<sup>4</sup> Der Softwareprototyp zur Erschließung des Berichts Umweltdaten 91/92 mit Hilfe des UBA-Thesaurus wurde von Margit Gaul implementiert. Eine detaillierte Beschreibung der Implementierung findet sich in (Gaul 1995).



**Abbildung 2:** Hypertextstruktur des mit dem Thesaurus verschlagworteten Berichts

wobei der Titel einer Hypertext-Seite durch den Titel des Unterkapitels gegeben ist. Abbildungen und Tabellen sind ebenfalls auf eigene Hypertext-Seiten ausgelagert und sind über Hyperlinks mit den Textseiten verknüpft, denen sie angehören. Das Inhaltsverzeichnis wird durch eine zweistufige Hierarchie von Hypertext-Seiten gebildet. Die Inhaltsverzeichnisse und die in ihnen aufgeführten Berichtsabschnitte sind untereinander bidirektional durch Hyperlinks verknüpft; außerdem besitzen die einzelnen Berichtsabschnitte Verzeigerung zu ihren Vorgängern und Nachfolgern im Sinne der linearen Ordnung der gedruckten Textfassung.<sup>5</sup>

### **Einbindung des Umwelt-Thesaurus**

Für die Erschließung des Inhalts von Umweltberichten wurde im FAW der Umwelt-Thesaurus des Umweltbundesamts (UBA) genutzt, der in Form einer Textdatei vorlag (Batschi 1994). Diese Textdatei enthält maschinell interpretier-

<sup>5</sup> Ein ähnliches, aber stärker automatisiertes Verfahren für die Konvertierung von Berichtstexten, die im Microsoft Word-Format und anderen Formaten vorliegen, wird in der Phase II des Projekts GLOBUS vom Forschungszentrum Karlsruhe, Institut für Angewandte Informatik zum Zweck der Bereitstellung von Berichten zur Altlastenthematik über WWW auf der Basis von Public Domain Software entwickelt.

bare Beschreibungen von über 17000 Begriffen und Synonymen aus dem UBA-Thesaurus. Bei den Synonymbegriffen ist eindeutig vermerkt, für welche Begriffe, bzw. für welche Kombination von Begriffen diese stehen. Bei den Begriffen sind die Namen der zugehörigen Oberbegriffe, Unterbegriffe und verwandten Begriffe sowie der für diesen Begriff benutzten Synonyme in eindeutig interpretierbarer Form angegeben.

Der Thesaurus wird für die Verschlagwortung der Berichtsabschnitte verwendet. Zu diesem Zweck wird in den einzelnen Berichtsabschnitten nach Thesaurus-Begriffen und nach deren Synonymen gesucht. Ein Begriff ist Kandidat für die Verschlagwortung eines Berichtsabschnitts, wenn er selbst oder eines seiner Synonyme im Text gefunden wurde. Da mitunter eine sehr große Anzahl von Thesaurus-Begriffen in den Texten gefunden wird, werden nur die sieben am häufigsten vorkommenden Begriffe übernommen und am Anfang des Berichtsabschnitts als Schlagwörter eingetragen. Mit Hilfe eines Editors ist es möglich, noch Änderungen (Hinzufügungen, Streichungen) an der Liste der ermittelten Schlagwörter vorzunehmen.

Bei der Suche nach Schlagwörtern wurde ein Verfahren benutzt, das insensitive gegenüber Groß- und Kleinschreibung ist und die Umlaute „ä“, „ö“ und „ü“ sowie „ß“ gleich behandelt wie die zweibuchstabigen Umschreibungen „ae“, „oe“, „ue“ und „ss“. Um das Verfahren gegenüber Beugungsendungen insensitive zu machen, wird ein Textvergleichsverfahren verwendet, das die letzten beiden Buchstaben der zu überprüfenden Wörter ignoriert. Die Entwicklung eines komplexeren Textanalyseprogramms war im Rahmen dieser Arbeit nicht beabsichtigt, da solche Programme bereits in anderen Retrievalsystemen existieren. Es ist aber jederzeit möglich, das verwendete einfache Schlagwortsuchprogramm durch ein leistungsfähigeres Textanalyseprogramm zu ersetzen.

Da jedes der Wörter eines Berichtsabschnittes mit den 17000 Begriffen des UBA-Thesaurus verglichen werden muß, war es erforderlich, einen sehr effizienten Algorithmus zur Schlagwortsuche zu realisieren. So wurde aus einem zu untersuchenden Berichtsabschnitt zunächst eine Wortliste erstellt und diese alphabetisch sortiert. Der Thesaurus selbst liegt von vornherein in einer alphabetisch sortierten Form vor. Daraufhin wurde der Vergleich der Begriffe aus beiden Listen im Rahmen eines Merge-Verfahrens durchgeführt. Dadurch beschränkt sich die Zahl der Vergleiche auf die Summe der Längen der beiden Listen. Da in der Regel der Thesaurus weitaus mehr Begriffe enthält als die Wortliste des Berichtsabschnitts, müssen für die Verschlagwortung eines Berichtsabschnitts nicht wesentlich mehr als 17000 Vergleiche durchgeführt werden, was auf einer schnellen Workstation im Sekundenbereich durchführbar ist.

Über die Verschlagwortung der Berichte hinaus wird der Thesaurus auch zur Erstellung eines Hypertexts genutzt, der die polyhierarchische Struktur der Thesaurus-Begriffe wiedergibt. Zu diesem Zweck wird für jeden Begriff aus dem Thesaurus eine Hypertext-Seite generiert. Diese Hypertext-Seite besitzt den Namen

des entsprechenden Begriffs als Titel. Unter der Überschrift BF („Benutzt für“) ist eine Liste der für diesen Begriff verwendeten Synonyme aufgeführt. Oberbegriffe, Unterbegriffe und verwandte Begriffe sind unter den Überschriften OB, UB und VB in Form von Hyperlinks aufgeführt, die auf die Hypertext-Seiten verweisen, die den jeweiligen Thesaurus-Begriff repräsentieren (Abbildung 2, rechter Teil).

Außerdem wird für alle Begriffe und Synonyme ein alphabetisches Verzeichnis – ebenfalls in Form von Hypertext – erzeugt. Bei den Synonymen ist hinter den Markierungen BS („Benutze Synonym“) bzw. BK („Benutze Kombination“) vermerkt, für welche Begriffe bzw. Begriffskombinationen sie stehen (siehe auch Abbildung 1, linkes Fenster). Über Hyperlinks kann von diesem Verzeichnis direkt in den Thesaurus zu der Hypertext-Seite gesprungen werden, die den jeweiligen Begriff darstellt.

Wenn der Benutzer der Verschlagwortungssoftware – ggf. nach einer manuellen Korrektur – mit der Zuordnung der Thesaurus-Begriffe zu den Berichtsabschnitten zufrieden ist, kann ein Programm gestartet werden, das den Eintrag der Berichtsabschnitte in die jeweils zugeordneten Begriffsdokumente vornimmt. Dazu werden in allen Begriffsdokumenten Hyperlinks auf die zugehörigen Berichtsabschnitte eingetragen, die den jeweiligen Thesaurusbegriff als Schlagwort besitzen. Nach diesem Schritt besitzen sowohl der Umweltbericht wie auch der Thesaurus eine geschlossene Form als Hypertext-Dokument und es bestehen bidirektionale Hyperlink-Verbindungen zwischen Berichtsabschnitten und den ihnen zugeordneten Thesaurus-Begriffen.

### **Volltextsuche**

Für die Volltextsuche wurde ein Public Domain Tool mit Namen *Swish* verwendet (Hughes 1995). Dieses Programm erlaubt es, einen Volltextindex über eine Menge von Textdokumenten aufzubauen. Ein Suchprogramm, das über ein Hypertextformular aufrufbar ist, nutzt den Volltextindex für die schnelle String-Suche.

### **Weiterführende Entwicklungen**

Mit der im vorigen Abschnitt beschriebenen, weitgehend noch im Rahmen der ersten Phase des GLOBUS-Projekts erstellten Software zur Unterstützung der thesaurusbasierten Recherche im WWW liegt bereits eine praktisch einsetzbare Lösung für die Gestaltung eines Berichtsservers auf dem WWW vor. Für die Archivierung von Berichten zur Altlastenthematik im Rahmen des GLOBUS-Teilprojekts AlfaWeb wird diese Software vom Forschungszentrum Karlsruhe bereits prototypisch eingesetzt (Weidemann 1996).

Darüber hinaus arbeitet das FAW an der Erweiterung der Funktionalität der Software. Ziel dieser Arbeiten ist es, stärker noch als in der bisherigen Lösung vielfältige Indexierungs- und Retrieval-Mechanismen zu unterstützen. Ein wesentlicher Schritt ist dabei die Ablage navigatorischer Metainformation in einer Datenbank. Dabei kann auf vorteilhafte Weise der Umstand genutzt werden, daß der Umwelt-

datenkatalog (UDK) ab Version 2.0 eine Datenbank umfaßt, die den Umweltthesaurus des Umweltbundesamts entsprechend dem relationalen Datenbankmodell zur Verfügung stellt (Niedersächsisches Umweltministerium 1994, 1995). Unter Berücksichtigung der weiteren Datenmodellierungen des UDK (insbesondere der UDK-Objekte) kann so ein Konzept zur Nutzung dieses Thesaurus zum Zugriff auf entsprechend verschlagwortete Umweltberichte entwickelt werden. Dabei werden insbesondere die folgenden Anforderungen berücksichtigt:

- die Ausweitung des thesaurus-unterstützten Zugriffs auf allgemeine verteilte Daten und Dienste (nicht nur Umweltberichte) und
- die Integration weiterer Indizierungs- und Recherchetechniken (insbesondere im Hinblick auf die Verwendung von Orts- und Zeitbezügen).

### **Ausdehnung der Techniken auf allgemeine verteilte Informationsquellen**

Berichte sind nur eine Art von Informationsquellen, die Auskunft über umweltrelevante Fragestellungen geben. Andere Informationsquellen sind Daten (z.B. aus Umweltdatenbanken oder Geoinformationssystemen) und allgemeine Methoden, die in Form von Dienstprogrammen realisiert sind. Im WWW wird eine integrierende Sicht auf alle diese Informationsquellen erschlossen, indem sie gleichermaßen in Form von Hypertext präsentiert werden. Es ist infolgedessen sinnvoll, die für die Berichte genutzten Indizierungstechniken auf alle diese Informationsquellen auszuweiten. Für die Repräsentation der Metainformation zu den verschiedenen Arten von Umweltinformationsquellen bietet sich dabei ein Klassenkonzept an, wie dies von O. Günther (1995, S. 43ff.) vorgeschlagen wird.

Informationsquellen liegen oft verteilt auf unterschiedlichen Servern vor. Um die Navigation zu diesen Informationsquellen wirkungsvoll zu unterstützen, bietet es sich an, die Metainformation über diese Informationsquellen auf einem zentralen Server zu verwalten. Dieser Metainformationsserver kann genutzt werden, um nach allen Arten von Informationsquellen zu suchen und um über diese eine erste Information zu erhalten. Dieser Server benötigt ein Netzwerkadressen-Verzeichnis, in dem für jede katalogisierte Informationsquelle eine Netzadresse (ein sogenannter *URL=Uniform Resource Locator*) vermerkt ist, unter dem diese Informationsquelle im WWW zu finden ist.

### **Fortgeschrittene Indizierungs- und Recherchetechniken**

Durch die Repräsentation des Thesaurus in der relationalen Datenbank sowie durch eine zusätzliche Relation, welche Begriffe aus dem Thesaurus und UDK-Objekte miteinander verknüpft, ist es möglich, vermittels Datenbankabfragen Umweltinformationen zu ermitteln, die mit bestimmten Begriffen verschlagwortet sind. Insbesondere ist es durch die Speicherung der Thesaurusinformation in der Datenbank auch möglich, Bezüge zu Schlagworten mit logischen Operatoren (*not*, *and*, *or*) zu verknüpfen. Durch die Begriffe aus dem UBA-Thesaurus und die logischen Operatoren der Datenbankabfragesprache ist damit ein Standardvokabular zur Formulierung von Fachbezügen festgelegt.

Weitere Erweiterungsmöglichkeiten der Schlagwortsuche bestehen in einer stärkeren Integration der Volltextsuche mit der thesaurusbasierten Suche sowie in der Unterstützung der mehrsprachlichen Nutzung mit Hilfe eines multilingualen Thesaurus.

Außer den Begriffen aus dem Umweltthesaurus und anderen Schlagwörtern können weitere Arten von Merkmalen zur Indizierung von Berichten und anderen Umweltinformationsquellen dienen. Hierzu zählen in erster Linie Zeitbezüge und Ortsbezüge. Wichtig ist aber auch hier, daß dadurch ein Standardvokabular festgelegt wird, das nicht nur zur Dokumentation der Informationen, sondern vor allem auch für die automatisierte Recherche genutzt werden kann. Im UDK ist ein solches Vokabular – mit Ausnahme des integrierten Umweltthesaurus – derzeit noch nicht gegeben. Das FAW untersucht deshalb Erweiterungen des UDK-Datenbankschemas, die für die maschinelle Recherche nach Umweltinformation bezüglich weiterer Kriterien, insbesondere Zeit- und Ortsbezügen, erforderlich sind.

So werden im UDK *Zeitbezüge* derzeit noch über Textfelder dargestellt. Im Hinblick auf ein standardisiertes Vokabular ist es erforderlich, von Textfeldern auf Datumsfelder überzugehen, die eine automatische Recherche nach Umweltinformationen über die Angabe von Zeitpunkten und Zeitintervallen erlauben. Ein Konverter zur teilautomatischen Umsetzung von informellen Zeitbezügen aus dem UDK in ein derartiges formalisiertes zeitliches Bezugssystem wird derzeit am FAW entwickelt.

Auch die Angabe von *Ortsbezügen* erfolgt im UDK derzeit noch über verschiedene Textfelder; wobei sowohl Koordinatenangaben als auch Angaben geographischer Gebiete möglich sind. Ein einheitliches, maschinell interpretierbares Format der Ortsangaben ist hierbei nicht festgelegt. Im Rahmen von GLOBUS II arbeitet das FAW an einem erweiterbaren Ortsverzeichnis, einem sogenannten *Gazetteer*, der für Ortsbezüge ein standardisiertes Vokabular bereitstellt. Solche Ortsbezüge, die jeweils aus einer geometrischen Beschreibung (im einfachsten Fall aus einem Koordinatenrechteck) und einem Namen bestehen, können den UDK-Objekten zugeordnet werden. Das bereits angeführte Konverterprogramm erlaubt eine teilautomatische Migration der prosaischen Ortsbezüge des UDK in das durch den Gazetteer bereitgestellte Standardvokabular.

Der Gazetteer kann in einem verallgemeinerten Sinn als Thesaurus angesehen werden. Die Ortsbezüge im Gazetteer sind in Bezug auf räumliches Enthaltensein ähnlich wie Thesaurusbegriffe polyhierarchisch geordnet. Hinzu kommt die Strukturierung der Ortsbezüge hinsichtlich räumlicher Überlappung sowie die geometrische Beschreibung des Ortsbezugs durch Koordinaten. Somit kann eine ortsbezogene Suche nach Umweltinformationen sowohl über die Angabe eines Suchrechtecks als auch durch Angabe geographischer Namen erfolgen. Durch die besondere Struktur des Gazetteer können bei einer solchen Suche auch Umweltinformationen mit übergeordnetem, untergeordnetem oder überlappenden Ortsbezug gefunden werden.

Eine wesentliche Bedeutung besitzt auch die Frage der Erfassung und Pflege der Metainformation in einem Metainformationsserver. Die einfachste, aber auch aufwendigste Methode ist die des manuellen Eintrags von Fach-, Orts- und Zeitbezügen. Hierbei verfolgt das FAW das Ziel, den Metainformationsserver mit einer WWW-Schnittstelle auszustatten, die den Anbietern der Umweltinformationen selbst die Eingabe der Metainformationen zu ermöglicht. Ein noch weitergehendes Ziel ist die automatische Extraktion von Metainformation aus den Quellinformationen. Die bereits beschriebene automatische Verschlagwortung von Umweltberichten mittels eines Thesaurus bzw. über einen Volltextindex ist hierfür ein Beispiel. In analoger Weise können auch mit Hilfe des Gazetteers aus Berichten und anderen textuellen Informationen geographische Namen extrahiert und Ortsbezüge hergeleitet werden. Schließlich können auch Zeitbezüge in vielen Fällen durch eine syntaktische Analyse aus den Textquellen erschlossen werden.

Weitere Anforderungen ergeben sich, wenn nicht nur textuelle Informationen, sondern auch Daten und Dienstprogramme zur Indizierung herangezogen werden sollen. Neben der Möglichkeit des manuellen Eintrags von Metainformationen durch den Datenhalter oder Dienstbetreiber kommen auch hier automatische Methoden in Betracht, etwa durch die Auswertung von Data Dictionaries oder durch eine intelligente Wissenserwerbskomponente im Sinne eines *Data Mining*.

### **Zusammenfassung und Ausblick**

Das Angebot von Umweltinformationen, die durch eine Vielfalt von Institutionen über WWW bereitgestellt werden, nimmt in beschleunigtem Maß zu. Das Auffinden relevanter Informationen in diesem komplexen Angebot ist bereits heute ein erhebliches Problem für Auskunftsuchende aus Politik, Verwaltung, Wirtschaft, Wissenschaft und Öffentlichkeit. Intelligente Techniken zur Unterstützung der Recherche nach diesen Informationen gewinnen deshalb zunehmend an Bedeutung.

Von Vorteil ist, daß die Informationen auf dem Internet in digitaler, maschinell interpretierbarer Form vorliegen. Dadurch werden automatische Indizierungs- und Suchtechniken anwendbar, die auf der Nutzung von Metawissen über Fach-, Zeit- und Ortsbezüge von Umweltinformationen beruhen. Metasystemen auf der Basis von Thesauren und Gazetteers kommt hierbei eine entscheidende Schlüsselrolle zu. Sie stellen eine entscheidende Ressource für ein effizienteres und wirtschaftlicheres Umweltinformationsmanagement dar.

## Literatur:

- Batschi, W.-D. (1994): Environmental Thesaurus and Classification of the Umweltbundesamt (Federal Environmental Agency), Berlin. In: Stancikova, P. / Dahlberg, I. (Hrsg.): Environmental Knowledge Organization and Information Management. INDEKS Verlag, Frankfurt/Main, S. 57–62.
- Gaul, M. (1995): Thesaurusgestützter Zugriff zu Umweltberichten in einem netzübergreifenden Hypertextsystem. Forschungsbericht FAW-TR-95006. FAW Ulm.
- Günther, O. (1995): Gutachten zur Entwicklung des Umwelt-Datenkatalogs (UDK). Erstellt im Auftrag des Niedersächsischen Umweltministeriums.
- Hughes, K. (1995): SWISH – Simple Web Indexing System for Humans. Enterprise Information Technologies (EIT), Menlo Park, California.  
URL: <http://www.eit.com/software/swish/swish.html>
- Niedersächsisches Umweltministerium / Bundesministerium für Umwelt, Jugend und Familie Republik Österreich / Dr. Lippke und Dr. Wagner GmbH (1994): UDK-T – Das Thesaurusmodul im Umwelt-Datenkatalog. Hannover / Wien.
- Niedersächsisches Umweltministerium / Bundesministerium für Umwelt, Jugend und Familie Republik Österreich (1995): Umwelt-Datenkatalog – Version 3.0 – Benutzerhandbuch. Hannover / Wien.
- Riekert, W.-F. (1995): Cooperative Management of Data and Services for Environmental Applications. In: Huber-Wäschle, F. / Schauer, H. / Widmayer, P. (Hrsg.): GIS 95 – Herausforderungen eines globalen Informationsverbundes für die Informatik. Informatik aktuell. Springer-Verlag, Berlin – Heidelberg – New York, S. 618–625.
- Riekert, W.-F. / Henning, I. / Schmidt, F. (1995) Integration von heterogenen Komponenten des Umweltinformationssystems (UIS) Baden-Württemberg. In: Güttler, R. / Geiger, W.: Integration von Umweltdaten, 2. Workshop, Schloß Dagstuhl 1994. Metropolis-Verlag, Marburg, S. 89–100.
- Umweltministerium Baden-Württemberg / Landesanstalt für Umweltschutz (1992): Umweltdaten 91/92.  
URL: <http://www.faw.uni-ulm.de:9876/Umweltdaten/start.html>
- Umweltministerium Baden-Württemberg / Landesanstalt für Umweltschutz / Forschungszentrum Informatik an der Universität Karlsruhe / Forschungsinstitut für anwendungsorientierte Wissensverarbeitung an der Universität Ulm / Forschungszentrum Karlsruhe – Technik und Umwelt / Institut für Photogrammetrie und Fernerkundung, Universität Karlsruhe / Institut für Kernenergetik und Energiesysteme Universität Stuttgart (1996): GLOBUS – Konzeption und prototypische Realisierung einer aktiven Auskunftskomponente für globale Umwelt-Sachdaten im Umweltinformationssystem Baden-Württemberg – Phase II. Abschlußbericht. Wissenschaftliche Berichte des Forschungszentrums Karlsruhe. Bericht FZKA 5700. Karlsruhe. URL: <http://www.iai.fzk.de/~weideman/doc/globus2/>
- Weidemann, R. / Geiger, W. / Reißfelder, M. (1996): AlfaWeb – Altlasten-Fachinformationen im World-Wide Web. Forschungszentrum Karlsruhe.  
URL: <http://www.iai.fzk.de/~weideman/lfu/lfu.htm>