EnviroInfo 2000: Umweltinformatik '00 Umweltinformation für Planung, Politik und Öffentlichkeit
Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

Georeferenzierung als Mittel zur Erschließung von Fachinformationen in Internet und Intranet

Wolf-Fritz Riekert[1] und Peter Treffler[2]

Abstract

Die Technik der Georeferenzierung eröffnet neuartige Methoden zur Erschließung von Fachinformationen im Internet und Intranet. Leistungsfähige Benutzungsoberflächen ermöglichen eine wirkungsvolle Informationsrecherche über geographische Namen oder über Zeigeaktionen mit der Maus und sind insbesondere im Umweltbereich von großem Nutzen. Dies wird am Beispiel der Weiterentwicklung des Geothesaurus des Umweltbundesamt Berlin demonstriert.

1. Ausgangspunkt

Die Zunahme von Informationen, die auf dem Internet wie auch in organisationseigenen Intranets 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 oder Intranet bereitstellen. 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 gefunden Hypertextseiten vornehmen. Da über

[1] Hochschule für Bibliotheks- und Informationswesen, Wolframstr. 34, D-70191 Stuttgart, Email: riekert@hbi-stuttgart.de
[2] Umweltbundesamt, Bismarckplatz 1, D-14139 Berlin, Email: peter.treffler@uba.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. Für diesen Zweck wurden Metainformationssysteme entwickelt, die eine Katalogisierung und Recherche von Informationen nach derartigen Kriterien erlauben. Bei der Konzeption von derartigen Metainformationssystemen hat sich gezeigt, dass es sinnvoll ist, insbesondere drei Arten von Beschreibungselementen für Informationen zu unterscheiden, nämlich *Zeitbezüge*, *Fachbezüge* und *Raumbezüge* (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. Beispiele für Metainformationssysteme, die derartige Suchkriterien unterstützen, finden sich im Umweltinformationsnetz Deutschland GEIN (German Environmental Information Network) (Bilo/Streuff 2000) und im Geographischen Informationssystem Umwelt GISU (Balzer/Nouhuys 1998) des Umweltbundesamts.

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

Schwieriger gestaltet sich die Behandlung von Fach- und Raumbezügen. Die einfachste und wirkungsvollste Möglichkeit, Fachbezüge festzulegen, geschieht mit Hilfe von Schlagwörtern. Wünschenswert ist allerdings ein kontrolliertes Vokabular für die Verwendung von Schlagwörtern. Dabei kann insbesondere auf Erfahrungen aus dem Dokumentationsbereich mit sogenannten Thesauri zurückgegriffen werden (Batschi 1994). Ein solcher Thesaurus besitzt zweierlei Funktionen. Zunächst stellt er einen Katalog dar, der die Namen aller Deskriptoren enthält, die für die Verschlagwortung von und die Recherche nach 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

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

semantischen Netz verknüpft sind. Im wesentlichen handelt es sich dabei um drei Beziehungstypen, nämlich die Synonymbeziehung, die Ober-/Unterbegriff-Beziehung und die Beziehung zwischen verwandten Begriffen. So lässt sich in einem solchen Thesaurus festlegen, dass „Südfrucht“ ein Oberbegriff des Begriffs „Orange“ und dieser wiederum ein Synonym des Begriffs „Apfelsine“ ist. Es bedarf keiner weiteren Erläuterung, dass der erste Teil der o.a. Recherche (Suche nach Informationen mit Fachbezug Südfrüchte) mit Hilfe des in einem Thesaurus repräsentierten begrifflichen Wissens durch ein intelligentes System automatisiert werden kann.

Die Interpretation des zweiten Teils der o.a. 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. Prinzipiell sind alle diese Möglichkeiten der Angabe von Raumbezügen sowohl bei der Recherche wie auch bei der Katalogisierung von Informationen möglich.

2. Gazetteers

Zur Verarbeitung von Raumbezügen ist geographisches Wissen erforderlich. Dieses lässt sich repräsentieren mit Hilfe eines sog. *Gazetteers*, auch *Geothesaurus* genannt. 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 müssen

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

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 Bodensee auch alle Informationen gefunden werden können, die mit der Insel Mainau assoziiert sind.

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 Vergröberung 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 durchnumerieren und über eine Tabelle mit den Regionen assoziieren. Diese Möglichkeit wird in GISU und in GEIN gewählt, da in diesen Systemen kein marktgängiges 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 und Katalogisierung von Informationen in einem Metainformationssystem genutzt werden:

- Mit Hilfe eines *Browsers* für topographische Beziehungen ist es möglich, Raumbezüge in textueller Form zusammen mit hierarchisch übergeordneten oder untergeordneten sowie überlappenden Raumbezügen darzustellen und zu diesen Raumbezügen zu navigieren. Die Namen der aufgesuchten Raumbezüge können bei der Recherche nach Informationen als Suchkriterien übernommen werden. Sie können auch genutzt werden für die manuelle Zuordnung von Raumbezügen zu Informationsressourcen für deren Katalogisierung im Metainformationssystem.
- Für eine Automatisierung von raumbezogenen Recherchen ist es sinnvoll, über Optionen festzulegen, inwieweit weitere Regionen aufgrund geometrisch/topologischer Beziehungen (Umschließung, Überlappung) mit in die Recherche aufgenommen werden sollen.
- Darüber hinaus können die geographischen Namen im Gazetteer auch verwendet werden, um eine *Textanalyse* von zu katalogisierenden Dokumenten hinsichtlich ihres Raumbezugs vorzunehmen. Eine textuelle Suche nach geographischen

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

Namen in einem zu katalogisierenden Dokument liefert die Kandidaten für die Georeferenzen, unter denen das Dokument im Metainformationssystem abgelegt werden soll.

- Herkömmliche *Suchmaschinen* können mit Hilfe der Information aus dem Gazetteer für die raumbezogene Recherche angepasst werden. Hierzu ist es möglich, einen Präprozessor für Benutzeranfragen zu entwickeln, der in einer Benutzeranfrage die Namen von Raumbezügen identifiziert und z.B. um weitere Namen von überlappenden Raumbezügen erweitert.
- Die geometrische Repräsentationsform von Raumbezügen ermöglicht auch *graphische Techniken* der Informationsrecherche und -katalogisierung in einem Raumbrowser. So lassen sich Raumbezüge durch direkte Manipulation mit Hilfe einer Maus spezifizieren. Alternativ ist eine Spezifikation von Raumbezügen durch Angabe von Koordinaten möglich.
- Ergebnisse von Anfragen an ein Metainformationsystem lassen sich auf einer kartographischen Oberfläche visualisieren. So kann eine Liste von gefundenen Informationsressourcen, welche jeweils einen Raumbezug besitzen, in Form sogenannter *Footprints* (Fußabdrücke) auf einer Landkarte dargestellt werden. Durch Anklicken eines solchen Footprints kann man dann auf die gewünschte Information direkt zugreifen.

Das hier vorgestellte Konzept, Informationen mit Raumbezügen zu versehen, ermöglicht neue Möglichkeiten für die Katalogisierung und Recherche von Informationen im 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, Informationsangebote auf dem Internet automatisch hinsichtlich ihres Raumbezugs zu indexieren.

3. Der Geothesaurus des Umweltbundesamts

Das im vorigen Abschnitt beschriebene Konzept steht und fällt mit den verfügbaren Inhalten eines solchen Gazetteers. Entscheidend für die Nutzbarkeit ist, dass ein reichhaltiges geographisches Vokabular angeboten wird, und dies ist nur möglich, wenn die Namen und Geometrien einer großen Zahl von geographischen Objekten aus allen relevanten Objektbereichen in den Gazetteer aufgenommen sind.

Aus diesem Grund wurde im Zuge des Aufbaus des Deutschen Umweltinformationsnetzes GEIN (German Environmental Information Network) und des Geographischen Informationssystems Umwelt GISU am Umweltbundesamt ein sehr reichhaltiger Gazetteer aufgebaut. In dem entstandenen System, dem *Geothesaurus des Umweltbundesamts*, sind weit über 50 000 administrative, physisch-geographische und fachliche Einheiten mit Namen und (punktförmigen, linearen oder flächenhaften) Geometrien erfasst.

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

Geometrisch-topologische Beziehungen zwischen den geographischen Einheiten im Geothesaurus des Umweltbundesamts werden über Rasterzellen hergestellt. Das

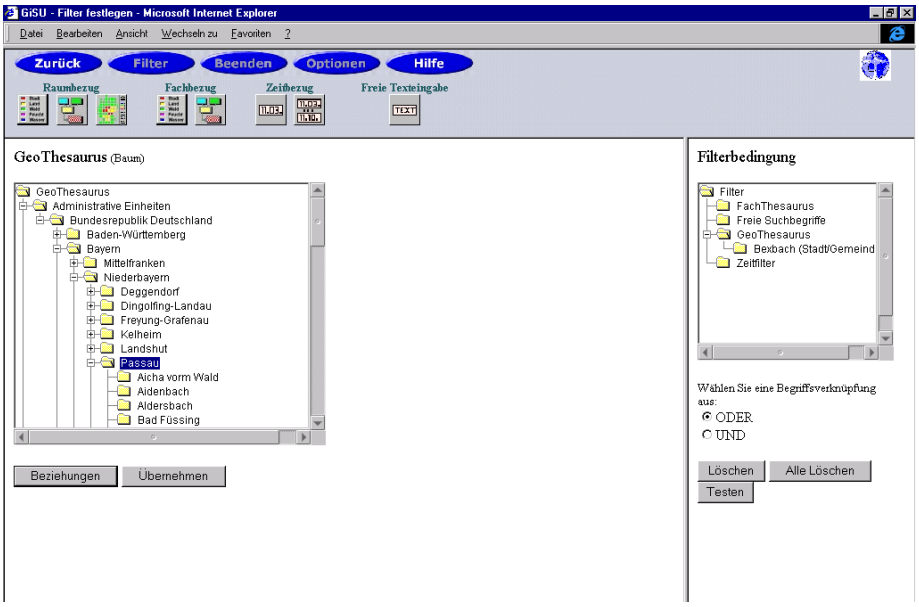


Abbildung 1: Raumfilter (systematischer Katalog)

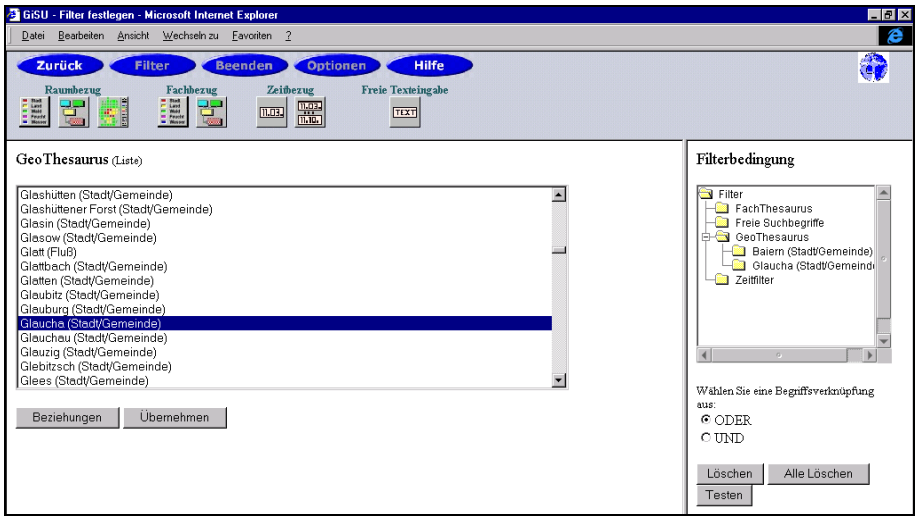
gesamte Plangebiet, das das Territorium der Bundesrepublik Deutschland umfasst, wurde in ein 3×3km²-Raster zerlegt und mit den Geometrien der geographischen Einheiten verschnitten. Als Ergebnis dieser Verschneidung (als das sogenannte „Spaltprodukt“) lassen sich geometrisch-topologische Beziehungen zwischen den geographischen Einheiten ableiten. Diese Beziehungen werden in den Systemen GEIN und GISU dazu genutzt, um die Indexierung und das Retrieval von Informationen mit Hilfe eines flexiblen geographischen Vokabulars zu unterstützen. Außerdem können im GISU-Raumbrowser durch Zeigeaktionen geographische Einheiten bezeichnet werden, die als Raumbezug für die Zwecke der Indexierung oder des Retrievals von Informationen genutzt werden können.

Abbildung 1 zeigt die Möglichkeit des Zugriffs auf Informationen mit Hilfe eines alphabetischen Katalogs der Namen räumlicher Einheiten. Im Unterschied dazu wird in Abbildung 2 ein systematischer Katalog der räumlichen Einheiten gezeigt. Abbildung 3 zeigt schließlich die Möglichkeit des Zugriffs aus Informationen über den GISU-Raumbrowser.

Die hier beschriebenen Entwicklungen bauen auf Konzepten und Prototypen der Firmen CadMAp und CONDAT sowie des Forschungsinstituts für anwendungsorientierte Wissensverarbeitung (FAW) Ulm auf und wurden nach Vorgaben aus dem Umweltbundesamt durch die Firmen Ernst Basler + Partner und

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

Umweltbundesamt durch die Firmen Ernst Basler + Partner und Sema Group realisiert.



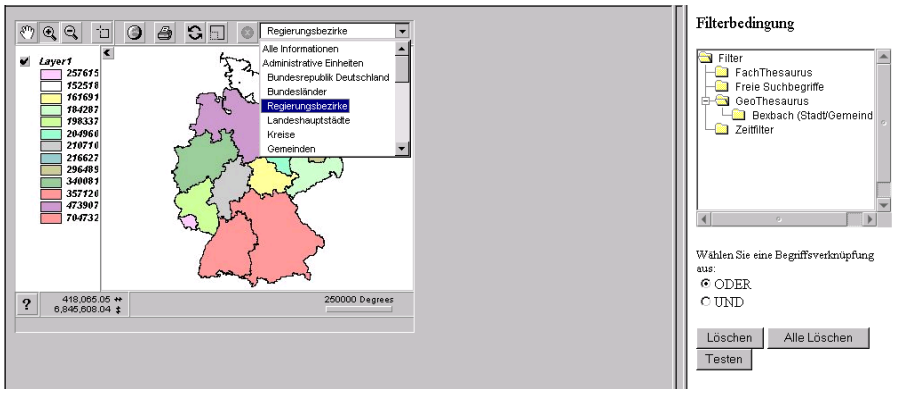


Abbildung 2: Zugriff über einen alphabetischer Katalog von räumlichen Einheiten

Abbildung 3: Zugriff über den Raumbrowser

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

4. Schlussfolgerungen

Die präsentierten Beispiele zeigen, dass der Einsatz von Gazetteers für die Verschlagwortung von und die Recherche nach Fachinformationen mit Metainformationssystemen und Suchmaschinen große Vorteile erbringt. Insbesondere die verbesserten Möglichkeiten der räumlichen Recherche sind vor allem im Umweltbereich von großem Nutzen. Solche Recherchen können wahlfrei durch Angabe geographischer Namen oder über Zeigeaktionen auf einer kartographischen Oberfläche spezifiziert werden.

Der Pflegeaufwand für diese Lösungen ist dabei vergleichsweise gering. Ähnlich wie in herkömmlichen Thesauri sind die Basisinformationen, die einem Gazetteer oder Geothesaurus zugrundeliegen, 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 Metainformationsysteme und Suchmaschinen genutzt werden. So ist der Geothesaurus des Umweltbundesamts bereits heute in zwei Systemen im Einsatz.

Deshalb zahlt sich die Arbeit, die in den Aufbau 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.

Literatur

Altavista (1998): http://www.altavista.com

Balzer, H., van Nouhuys, J. (1998): GISU – Geographisches Informationssystem Umwelt im Umweltbundesamt. In: Riekert, W.-F., Tochtermann, K. (Hrsg.): Hypermedia im Umweltschutz. 1. Workshop, Ulm 1998. Metropolis-Verlag, Marburg.

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.

Bilo, M. und Streuff, H. (2000): Das Umweltinformationsnetz Deutschland GEIN2000 – Fachliche Anforderungen an ein Forschungs- und Entwicklungsvorhaben. In: Tochtermann, K., Riekert, W.-F. (Hrsg.): Hypermedia im Umweltschutz. 3. Workshop, Ulm 2000. Metropolis-Verlag, Marburg.

Riekert, W.-F. et al. (1999): Erschließung von Fachinformationen im Internet mit Hilfe von Thesauri und Gazetteers. In: Dade, C., Schulz, B. (Hrsg.): Management von Umweltinformationen in vernetzten Umgebungen. 2. Workshop des Arbeitskreises „Hypermedia im Umweltschutz“ und 8. Workshop der Fachgruppe „Betriebliche Umweltinformationssysteme“, Nürnberg 1999. Metropolis-Verlag, Marburg.

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5

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.

Copyright 2000 Metropolis Verlag, Marburg, ISBN: 3-89518-307-5