

## Management von Umweltinformationen in vernetzten Umgebungen

---

2. Workshop des Arbeitskreises „Hypermedia im Umweltschutz“,  
8. Workshop der Fachgruppe „Betriebliche Umweltinformationssysteme“  
der Gesellschaft für Informatik, 25.-26. März 1999, Nürnberg

### Erschließung von Fachinformationen im Internet mit Hilfe von Thesauri und Gazetteers

Wolf-Fritz Riekert<sup>1,2</sup>, Christian Fuchs<sup>2</sup> und Gerd Klingler<sup>2</sup>

<sup>1</sup>Hochschule für Bibliotheks- und Informationswesen (HBI)  
Wolframstr. 34, D-70191 Stuttgart, Email: riekert@hbi-stuttgart.de

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



## Informationsangebote auf dem Internet

---

Informationsangebote im Internet / WWW  
(Informationsressourcen):

- multimediale Dokumente
- Daten
- Dienstprogramme

Angebot explosiv wachsend

- Problem: Orientierung im WWW (“Lost in Hyperspace”)
- Werkzeuge zur Informationssuche erforderlich



## Suchmaschinen

---

- Volltextindex über gesamtes WWW,  
aufgebaut und gepflegt durch sogenannte “Robots”
- Praktisch kein Aufwand für das Annoncieren von  
Datenbeständen
- Stichwortsuche über WWW-Browser

Aber:

- ausschließlich textuelle Auswertung der Stichwörter
- keine inhaltliche Interpretation
- funktioniert nur für in Textform vorliegende Information

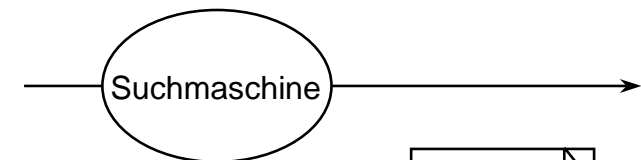


## Beispiel

---

Anfrage:

“Südfrüchte  
Bodensee”



## Metainformationssysteme

Metainformationssysteme ermöglichen Katalogisierung und Recherche von Informationen nach semantischen Kriterien:

- Fachbezug (z.B. "Südfrucht")
- Raumbezug (z.B. "Bodensee")
- Zeitbezug (z.B. "1995-1997")

Katalogisierung (= Eintrag der Metainformationen) erfolgt manuell durch Systemverwalter oder Informationsanbieter:

- Höhere Qualität
- Aber auch höherer Arbeitsaufwand



## Der Information Locator des GEIN-Prototyps: Beispiel eines Metainformationssystems

GEIN: German Environmental Information Network. Arbeiten am FAW im Auftrag des UBA (1996 und 1998)

## Formulierung und Verarbeitung von Fach- und Raumbezügen mit Thesauri und „Gazetteers“

### Anforderungen

- Vokabular zur Formulierung des Fachbezugs von Informationsressourcen
- Techniken zur automatisierten Verarbeitung von Fach- und Raumbezügen (⇒ Arbeitserleichterung)

### Lösungsansatz:

- Einsatz eines Thesaurus für die Formulierung und automatisierte Verarbeitung des Fachbezugs
- analog: Einsatz eines „Gazetteers“ für Raumbezug

**Anmerkung:** Die Verarbeitung des Zeitbezugs von Informationsressourcen ist kein Gegenstand dieses Vortrags



## Thesaurus

- stellt ein kontrolliertes Vokabular zur Festlegung von Fachbezügen bereit
- besteht aus einer strukturierten Sammlung von Begriffen
- Begriffe dienen zur Verschlagwortung und Recherche
- Begriffe sind dabei mehr als bloße Schlagwörter
- Begriffe bilden semantisches Netz, verknüpft durch:
  - ⇒ Synonymbeziehung (*Apfelsine - Orange*)
  - ⇒ Ober-Unterbegriff-Beziehung (*Südfrucht - Orange*)
  - ⇒ Verwandte Begriffe (*Frucht - Samen*)



## Problem der Blackboxsuche: Abhilfe durch Thesaurus

**Problem:** In den meisten Metainformationssystemen erfolgt die Suche nach Informationsressourcen über ein Formular („Blackboxsuche“)

- Erforderlicher Detaillierungsgrad der Anfrage unklar
- Ablehnung vor allem durch nichtgeübte Benutzer

**Anforderung:** Gewünscht wird ein Zugang zu den Informationen über hierarchische Verzeichnisse

- Aber: Solche Verzeichnisse sind manuell kaum pflegbar

**Lösungsansatz:** Automatische Generierung eines hierarchischen Verzeichnisses aus einem Thesaurus



## Generierung eines Verzeichnisses aus einem Thesaurus

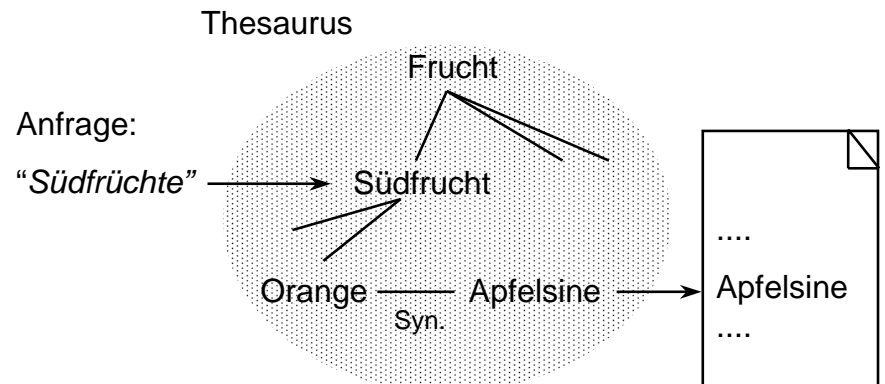


## Verfahren zur Generierung eines Verzeichnisses aus einem Thesaurus

- Erzeugung eines „ausgedünnten“ Thesaurus, bestehend aus den relevanten Begriffen, d.h.
  - ⇒ allen Thesaurusbegriffen, die zur Verschlagwortung von Informationsressourcen genutzt wurden
  - ⇒ plus allen Oberbegriffen, Ober-Oberbegriffen usw.
- Hierarchische Darstellung des Thesaurus in einem „Explorer“-Fenster, ausgehend von „Toplevel-Begriffen“
- Graphische Hervorhebung von Thesaurusbegriffen,
  - ⇒ die direkt zu Treffern führen,
  - ⇒ deren Unterbegriffe zu Treffern führen,
  - ⇒ bzw. für die beides gilt



## Rechercheunterstützung durch Thesaurus



## Kombination der Vorteile von Suchmaschinen und Metainformationssystemen durch Thesauri

Nutzung eines Metainformationssystems

- Komfortable Recherche nach semantischen Kriterien
- Aber auch: Hoher Aufwand für die Katalogisierung

Nutzung einer Suchmaschine

- Praktisch kein Aufwand für die Katalogisierung
- Aber Recherche nur nach textuellen Kriterien

Synthese

- Kombination der Vorteile von Suchmaschinen und Metainformationssystemen durch thesaurusbasierte Recherche mit Hilfe von Suchmaschinen



## Kombination der Vorteile von Suchmaschinen und Metainformationssystemen durch Thesauri

	Geringer Aufwand bei der Katalogisierung	Recherche nach semantischen Kriterien
Suchmaschine	X	
Metainformationssystem		X
Thesaurusbasierte Recherche mit Suchmaschine	X	X



## Thesaurusbasierte Recherche mit Hilfe einer Suchmaschine

The screenshot shows a web browser window titled 'Thesaurusbasierte Recherche: Kraftstoff - Microsoft Internet Explorer'. The page content includes a search bar, a list of synonyms (Autotreibstoff, Ottokraftstoff, Alkoholkraftstoff, Alkoholmotor), and a hierarchical tree structure of terms. Annotations with callout boxes point to specific features: 'Synonyme' points to the synonym list; 'Oberbegriffe' points to the 'Kraftstoff' node in the tree; 'Schwesterbegriffe' points to 'Brennstoff (fest)'; 'Übersetzung des selekt. Begriffs' points to the 'motor fuel' translation; 'Begriffshierarchie' points to the tree structure; 'Voreinstellungen' points to a settings panel on the right with options like 'Oberbegriffe', 'Unterbegriffe', 'Schwesterbegriffe', etc.; and 'Anfrage an Suchmaschine' points to the search input field at the bottom.



## Probleme bei der Formulierung des Raumbezugs in herkömmlichen Systemen

- Sehr starres Vokabular
  - ⇒ I.d.R. wird nur ein einziges Raumbezugssystem unterstützt (nur Koordinaten, nur Namen)
- Keine Intelligenz
  - ⇒ Es wird nicht erkannt, daß eine Region eine andere subsumiert
- Ausweg: Festlegung des Raumbezugs durch Geobjekte
  - ⇒ Diese besitzen Namen und Geometrie
  - ⇒ sowie geometrisch-topologische Beziehungen
  - ⇒ und bilden eine Geodatenbasis, einen sogenannten Gazetteer (Ortsverzeichnis)



## Gazetteer

- Geodatenbasis zur Spezifikation von Raumbezügen von Informationsressourcen
- Raumbezug: n:m-Beziehung zwischen Informationsressourcen und Gazetteer-Objekten
- Gazetteer-Objekte sind charakterisiert durch:
  - ⇒ Typ (z.B. "Gemeinde")
  - ⇒ Name (z.B. "Nürnberg")
  - ⇒ Identifikator (z.B. Verwaltungskennziffer)
  - ⇒ Geometrie (z.B. Polygon)
  - ⇒ ggf. Hierarchie (z.B. Verwaltungshierarchie)



## Topologisch-geometrische Beziehungen

Topologisch-geometrische Beziehungen zwischen Gazetteer-Objekten

- Überlappung
- Ineinander-Enthaltensein
- Berührung

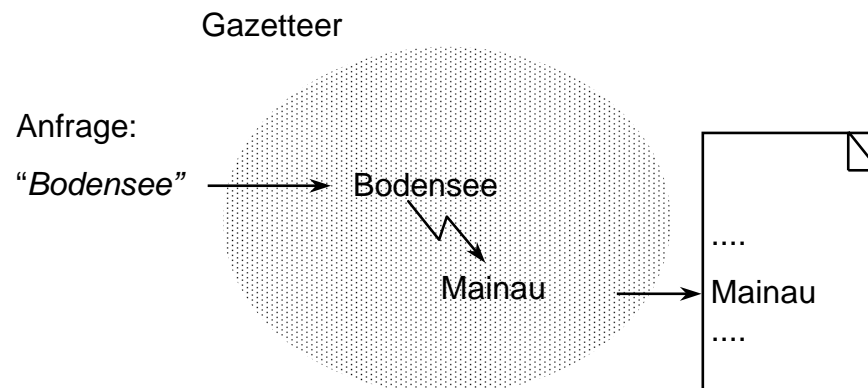
ableitbar aus Geometrie. Können genutzt werden für

- Katalogisierung und
- Recherche

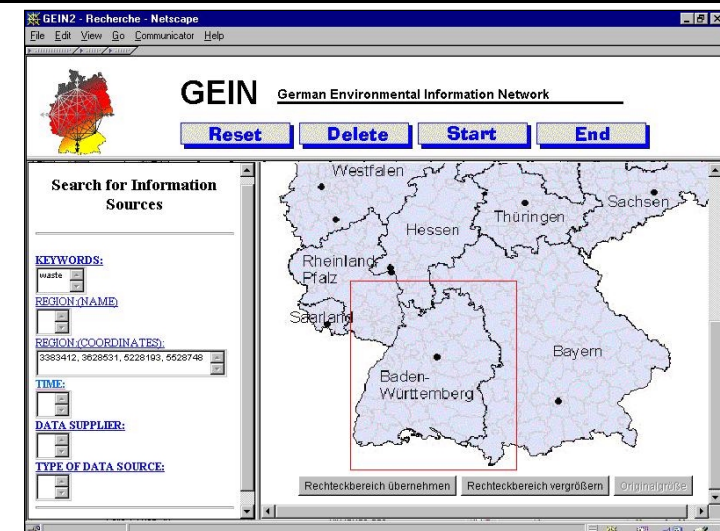
mit flexiblem geographischem Vokabular



## Rechercheunterstützung durch Gazetteer



## Graphische Selektion von Raumbezügen mit Hilfe des GEIN-Gazetteers



## Textuelle Spezifikation von Raumbezügen mit Hilfe des GEIN-Gazetteers

The screenshot shows the GEIN-Gazetteer interface in a Netscape browser window. The title bar reads "GEIN2 - Recherche - Netscape". The main content area is titled "GEIN German Environmental Information Network". Below the title are buttons for "Reset", "Delete", "Start", and "End". On the left, there is a "Search for Information Sources" section with fields for "KEYWORDS:" (containing "waste"), "REGION (NAME)", "REGION (COORDINATES)", "TIME", "DATA SUPPLIER", and "TYPE OF DATA SOURCE". The main search results area shows a list of regions under the heading "ACCEPT REGION AS ATTRIBUTE": "Baden-Baden (Stadtkreis)", "Baden-Württemberg", and "Wiesbaden (Kreisfreie Stadt)". Below this list are radio buttons for "Choose your Option for Search Region": "Without hierarchical search", "Use enclosing region", "Use subregion" (selected), and "Use enclosing- and subregion". At the bottom of the results area are buttons for "ACCEPT REGION AS ATTRIBUTE" and "BROWSE REGION". Callout boxes point to these elements with the following text: "Ergebnis der Suche nach „baden“", "Eingeschlossene Regionen übernehmen", "Selektierte Region übernehmen", and "Einstieg in Regionenbrowser". The FAW logo is visible in the bottom right corner of the screenshot.

## Gazetteer: Anwendungen

- Browser für geographische Namen
- Graphische Techniken zur Spezifikation von Raumbezügen
- Automatische Erweiterung von Suchanfragen durch Einschluß von Informationsressourcen mit engerem oder überlappendem Raumbezug
- Freier Wechsel zwischen geometrischer und textueller Repräsentation von Raumbezügen
- Textanalyse zur automatisierten Katalogisierung von Informationen
- Graphische Anzeige von Raumbezügen als "Footprints"



## Ergebnisse

- Höherer Nutzen von Metainformationssystemen und Suchmaschinen durch Thesauri und Gazetteers
- Neue, attraktive Benutzungsoberflächen: Landkarten, Verzeichnisse, Netzdarstellungen statt Blackboxsuche
- Flexibles Vokabular zur Spezifikation von Fach- und Raumbezügen
- Automatische Umformulierung, Erweiterung und Übersetzung von Schlagwörtern und Ortsangaben
- Thesauri und Gazetteers sind relativ stabile, weitgehende anwendungsunabhängig nutzbare Wissensstrukturen:
  - ⇒ Einfacher Pflegeaufwand, mehrfacher Nutzen

