

Georeferenzierung als Mittel zur Erschließung von Fachinformationen in Internet und Intranet

Wolf-Fritz Riekert
Hochschule für Bibliotheks- und
Informationswesen (HBI) Stuttgart

Peter Treffler
Umweltbundesamt
Berlin



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

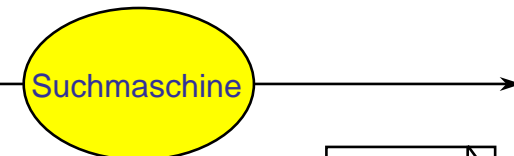
Aber:

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

Beispiel

Anfrage:

"Südfrüchte
Bodensee"



....
Apfelsine
....
Mainau

Metainformationssysteme

Metainformationssysteme ermöglichen Katalogisierung und Recherche von Informationen nach semantischen Kriterien:

- Fachbezug (z.B. "Südfrucht")
- Raumbezug (z.B. "Bodensee")
- Zeitbezug (z.B. "1998 – 2000")

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

- Höhere Qualität
- Aber auch höherer Arbeitsaufwand

Beispiele für Metainformationssysteme im Umweltbereich

- Umweltdatenkatalog
- Environmental Catalog of Data Sources (CDS)
- National Geospatial Clearinghouse (NGSC)
- InGeo IC (InGeoForum)
- GEIN
- GISU - Metakomponente

Formulierung und Verarbeitung von Fach-, Zeit- und Raumbezügen

Anforderungen

- Vokabular zur Formulierung des Fach-, Raum- und Zeitbezugs von Informationsressourcen
- Techniken zur automatisierten Verarbeitung von Fach-, Raum- und Zeitbezügen (⇒ Arbeitserleichterung)

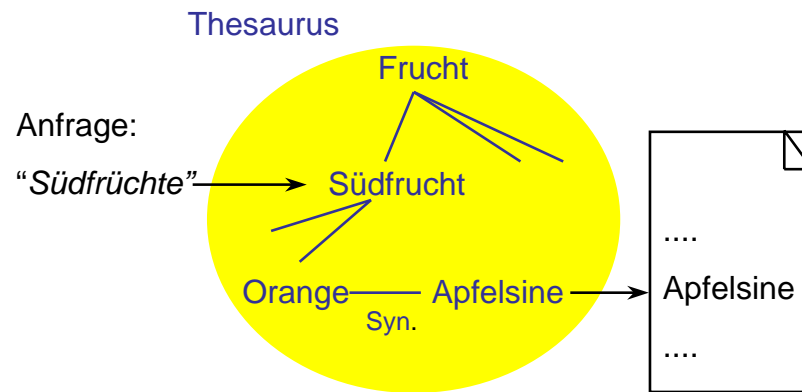
Lösungsansatz

- Einsatz eines Thesaurus für die Formulierung und automatisierte Verarbeitung des Fachbezugs
- analog: Einsatz eines sogenannten „Gazetteers“ für Raumbezug: Hauptgegenstand dieses Vortrags
- Zeitbezug: Behandlung vergleichsweise einfach, 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)

Rechercheunterstützung durch Thesaurus



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 Geoobjekte
 - ⇒ 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. "Bonn")
 - ⇒ Identifikator (z.B. Verwaltungskennziffer)
 - ⇒ Geometrie (z.B. Polygon)
 - ⇒ ggf. Hierarchie (z.B. Verwaltungshierarchie)
 - ⇒ ggf. Synonymnamen, fremdsprachliche Entsprechungen

Geometriedarstellung im Gazetteer

Alternative Realisierungen:

- Umschließendes Rechteck
- Bedeckte Rasterzellen
 - ⇒ SQL-Datenbank reicht aus für Realisierung
- Exakte Vektorgeometrie
 - ⇒ GIS erforderlich

Topologisch-geometrische Beziehungen

Topologisch-geometrische Beziehungen zwischen Gazetteer-Objekten

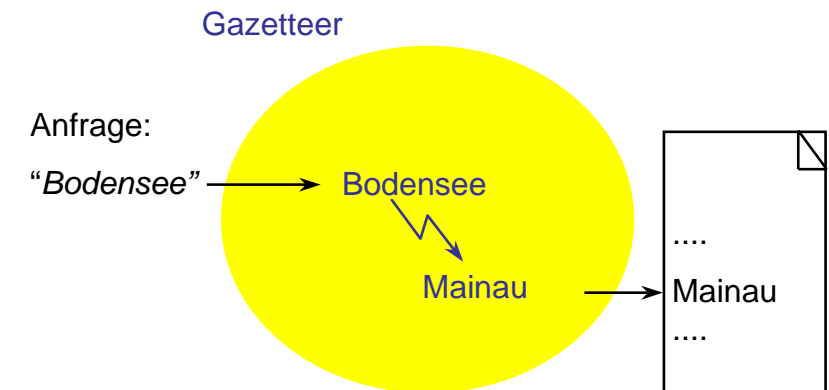
- Überlappung
- Ineinander-Enthaltensein
- Berührung

ableitbar aus Geometrie. Können genutzt werden für

- Katalogisierung und
- Recherche

mit flexiblem geographischem Vokabular

Rechercheunterstützung durch Gazetteer



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"

Umsetzung im Umweltbundesamt

Die hier beschriebenen Prinzipien wurden in zwei Systemen des Umweltbundesamts umgesetzt:

- German Environmental Information Network (GEIN)
- Geographisches Informationssystem Umwelt (GISU)

Konzept- und Prototypentwicklung:

- Forschungsinstitut für anwendungsorientierte Wissensverarbeitung (FAW), Ulm
- Condat AG, Berlin (nach Übernahme von Projektaktivitäten der CADMAP GmbH, Berlin)

Realisierung jetzt durch:

- Ernst Basler + Partner (GISU)
- Sema Group (GEIN)

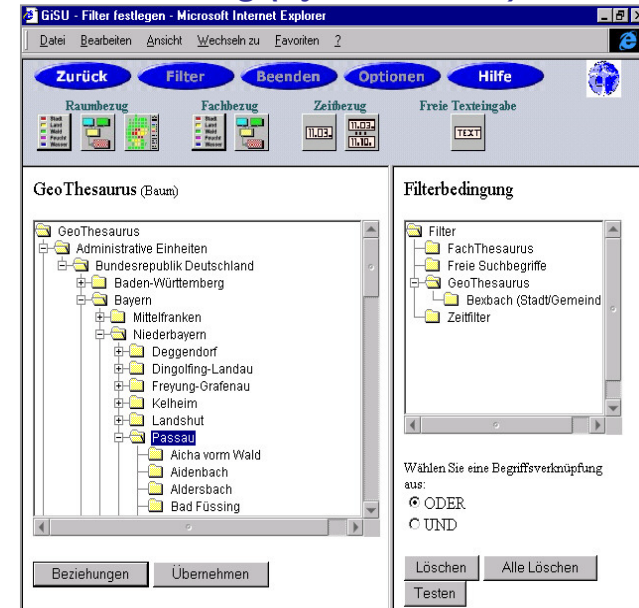
Der Geothesaurs des Umweltbundesamts

- Realisierung eines Gazetteers, genannt Geothesaurs
- enthält rund 100 000 administrative, physisch-geographische und fachliche Einheiten in Form von Geoobjekten
- Verschneidung der Geometrien mit einem 3x3 km²-Raster
 - ⇒ sogenanntes Spaltprodukt
 - ⇒ Geothesaurs kann in relationaler Datenbank repräsentiert werden, kein GIS erforderlich
- Geometrisch-topologische Beziehungen werden genutzt, um Indexierung und Retrieval in den Systemen GISU und GEIN zu unterstützen

Georeferenzierung

© Riekert/Treffler, 4.-6.10.2000 S. 17

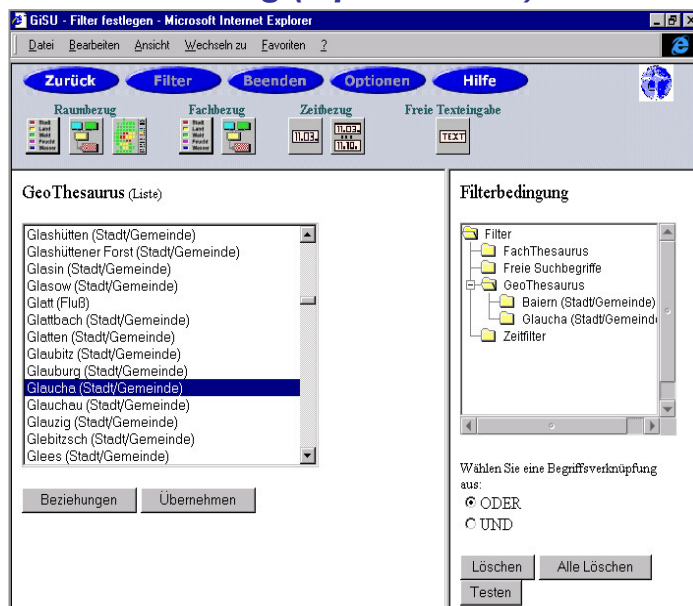
GISU: Raumbezug (systematisch) als Filter



Georeferenzierung

© Riekert/Treffler, 4.-6.10.2000 S. 18

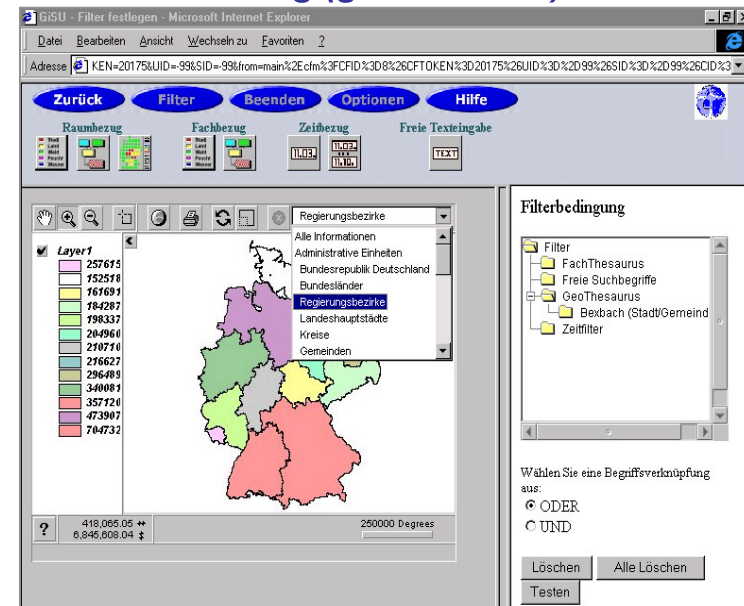
GISU: Raumbezug (alphabetisch) als Filter



Georeferenzierung

© Riekert/Treffler, 4.-6.10.2000 S. 19

GISU: Raumbezug (geometrisch) als Filter



Georeferenzierung

© Riekert/Treffler, 4.-6.10.2000 S. 20

Ergebnisse

- Höherer Nutzen von Metainformationssystemen und Suchmaschinen durch Einsatz eines Gazetteers
- Landkarten als attraktive Benutzungsoberflächen
- Flexibles Vokabular zur Spezifikation von Raumbezügen (analog zur Spezifikation von Fachbezügen mit Thesauri)
- Automatische Umformulierung, Erweiterung und Übersetzung von Ortsangaben
- Gazetteers sind (ähnlich wie Thesauri) relativ stabile, weitgehende anwendungsunabhängig nutzbare Wissensstrukturen:
 - ⇒ Einfacher Pflegeaufwand, mehrfacher Nutzen