

Zum Rechercheerfolg bei Metainformationssystemen

Erich Weihs¹

Kurzfassung

Im Bereich der Umweltvorsorge sind sprachlich basierte Ordnungssysteme von erheblicher Bedeutung: Die Erschließung und Nutzung vorhandener Daten für Umweltfragen wird zunehmend erkannt und umgesetzt (u.a. UDK, UOK, GEIN).

Da die meisten Daten über die Umwelt zeit- und raumbezogen vorliegen, sind neben den sachbezogenen Recherchen auch zeit- und raumbezogene Recherchen über Ortsnamen, bestimmte Raumeinheiten, topographische Begriffe oder Koordinaten erforderlich und bei Klassifikationssystemen zu berücksichtigen.

Auf diesem Wege ist es möglich, sowohl über textbasierte Ordnungsbegriffe wie über Koordinaten eines beliebig definierten Gebietes raumbezogen zu recherchieren. Der raumbezogene Zugang ist erforderlich, da die meisten Informationen auf der Basis von Flächenstrukturen (z.B. naturräumliche Einheiten wie Flusstäler, Gewässer, aquatische und terrestrische Bezeichnungen oder Ortsnamen) auszuwerten sind.

Die Existenz einer Vielzahl von Homonymen, also *einer* Bezeichnung für unterschiedliche Objekte, die u.a. aus *Sicht* der Anwendung unterschiedlicher Fachterminologien resultieren, ist besonders im Umweltbereich häufig. Im Gegensatz zu anderen Wissensbereichen existiert hier grundsätzlich kein einheitlicher Sprachgebrauch der eine allgemein gültige Beschreibung *und Einordnung* von Umwelteinflüssen ermöglichen würde. Im Gegenteil, Umweltfragen sind Querschnittsfragen, die viele etablierte Wissensbereiche treffen. Jeder Wissens- oder Fachbereich wird nach *seinem* Aspekt aus Sicht seines *Realitätsmodells* kategorisieren.

Grundlage für die Indexierung und Recherche sind neben Freitext auch Thesauren und Einordnungen nach verschiedenen Klassifikationssystematiken. Der gesicherte Rechercheerfolg hängt von der Nutzung der zutreffenden Fachsprache, der Klassifikationssysteme und Thesauren und/oder der geeigneten Umsetzung umgangssprachlicher Formulierungen ab.

In der Regel ist davon auszugehen, dass der Rechercheerfolg neben dem – nicht immer vorauszusetzenden kundigen Sprachgebrauch des Anwenders - von der Qua-

¹ Bayerisches Staatsministerium für Landentwicklung und Umweltfragen, München
erich.weihs@stmlu.bayern.de

lität einer vorhandenen - oft teilautomatisierten - Beschlagwortung in Verbindung mit dem Thesaurus wie den angewandten Klassifikationsverfahren abhängt. Entsprechend wird die Behandlung des Raum- und Zeitbezugs wie die von Homonymen unterschiedlich sein.

In der Literatur wird davon ausgegangen, dass die Trefferwahrscheinlichkeit bei einschlägigen Recherchen zwischen 50 bis 80 % liegt. Das bedeutet, dass trotz einer möglichen negativen Recherche die Wahrscheinlichkeit zwischen 20 und 50 % liegt, dass gesuchte Daten in der Grundgesamtheit vorhanden sind.

Da gerade Metainformationssysteme der Öffentlichkeit den Nachweis und die Recherche von Umweltinformationen ermöglichen sollen, ist hier der Indexierung und den Recherchemöglichkeiten besondere Aufmerksamkeit zu schenken.

In dem geplanten Beitrag wird auf Grundlage mathematisch-statistischer Methoden versucht die Abhängigkeit zwischen Rechercheerfolg und Freitextsuche versus thesaurusbasierter Suche auf Basis einer repräsentativen Fallzahl quantitativ darzustellen. Dabei werden auch umgangssprachliche Fragestellungen, wie sie vom fachunkundigem Benutzer gestellt werden, eine Rolle spielen. Grundgesamtheit der Untersuchung sind die Metainformationen aus GEIN, dem UDK und UOK unter Einbeziehung des sogenannten festen und datenbankgestützten Angebotes.

Literatur:

GEIN (1999, 2000)

UDK (1998, 1999)

Weih, E (1996): Zur Klassifikation umweltbezogener Daten im Rahmen des Bayerischen Umweltinformationssystems an Hand eines objektorientierten Ansatzes, in: nfd Zeitschrift für Informationswissenschaft und -praxis, Journal of Information Theory and Work 47, Heft 6/1996, 361-365, Weinheim

Weih, E. (1998a): Der Objektdatenkatalog des Bayerischen Informationssystems für Planung und Umwelt (ISPU) in: Umweltinformatik '98, Vernetzte Strukturen in Informatik, Umwelt und Wirtschaft, H..D. Haasis, K.C. Ranze (Hrsg./Eds.) Bd. II, 621 - 635

Weih, E. (1998b): On the classification of environmental data in the Bavarian Environmental Information System using an object-oriented approach in: Studies in Classification, data Analysis, and Knowledge Organization: Data Science, Classification, and Related Methods; Tokyo, 728 - 735

Weih.E (1999): Objektorientierung bei der Speicherung und Recherche von XML-Daten in vernetzten Systemen in: Management von Umweltinformationen in vernetzten Umgebungen, Bd. 21 Umweltinformatik aktuell, p. 58 - 69

Weih,E. (2000): Anwendung des XML Thesaurus im Umweltobjektkatalog, Beitrag zur Jahrestagung der Gesellschaft für Klassifikation, Passau 2000