

Inhalts- und strukturspezifische Such- und Navigationsinstrumente im Konstanzer-Hypertext-System (KHS)¹

Rainer Hammwöhner / Marc Rittberger / Volker Zink

Universität Konstanz

Informationswissenschaft

Zusammenfassung

Hypertextsysteme, die nicht als geschlossene Informationswelt angeboten werden, sondern ständigen Veränderungen in Struktur, Organisation und Umfang ausgesetzt sind, erfordern spezielle Such- und Navigationsinstrumente, die diesen Veränderungen Rechnung tragen. Das Konstanzer-Hypertext-System bietet dem Benutzer eine Vielzahl solcher Instrumente an. Zunächst wird in dem Beitrag auf das Hypertextmodell mit seinen typisierten Hypertextobjekten eingegangen und die verschiedenen Tools zur Implementation generischer Interaktionsformen werden vorgestellt. Danach beschreiben wir die Navigations- und Suchinstrumente des Konstanzer-Hypertext-Systems, sowie die Einbindung eines Thesaurus bei Navigation oder Suche.

Keywords: Browsing, Hypertextsystem, Hypertext-Modell, Information-Retrieval, KHS, Mensch-Maschine-Interaktion

Inhalt

1. Einleitung
2. Das Hypertext-Modell des KHS
3. Tools zur Implementation generischer Interaktionsformen
4. Navigation
5. Such- und Retrievalinstrumente im KHS
6. Einsatz von Thesauri in Suche und Navigation
7. Ausblick Literaturverzeichnis

I Einleitung

Im folgenden werden exemplarisch Tools des Konstanzer-Hypertext-Systems (im folgenden KHS) vorgestellt, die in ihrem Zusammenwirken eine Integration von Exploration und Suche, sowie von internem und externem Informationszugriff in ein weitgehend uniformes Interaktionsmodell gewährleisten.

Hypertext bietet in erster Linie Navigations- und Browsinginstrumente zur Informationssuche, die eine explorative Form der Interaktion unterstützen. Die Notwendigkeit weiterer Interaktionshilfen, wie sie beispielsweise im Information Retrieval üblich sind, erscheint bei offenen Hypertextsystemen, wie KHS, offensichtlich, da die rein explorative Interaktion bei Hypertexten mit einem sehr großen, ständig wachsenden Informationsangebot zur Selektion,

¹ Dieser Text ist erschienen in: Josef Herget (Hrsg.) Neue Dimensionen der Informationsverarbeitung, Proceedings des 1. Konstanzer Informationswissenschaftlichen Kolloquiums (KIK '93), UVK, 1993, S. 96-110.



Dieser Text ist unter der folgenden Creative Commons Lizenz lizenziert: Attribution-NonCommercial-NoDerivs 2.0 Germany (<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>).

Identifikation und Relevanzbeurteilung nicht mehr genügt [HR93]. Für sich alleine genommen können also weder traditionelle Verfahren des Information Retrieval noch hypertextspezifisches Browsen oder Navigieren die Anforderungen erfüllen, die an große, im voraus geordnete Informationsmengen (wie sie Hypertexte darstellen) gestellt werden. Werden beide Interaktionsstile zur Suche und Navigation in einem Hypertext angeboten, so sollten sie gemeinsam, kombiniert, also nicht unabhängig voneinander, zur Klärung des Informationsproblems des Benutzers dienen. Agosti [Ago88] stellt diese Kombination als ein neues Modell für das Information Retrieval dar, man sollte aber eher von einer Ergänzung und Integration von Information Retrieval und Hypertext sprechen, da die Such- und Navigationsinstrumente "nicht nur als Alternativen beim Einstieg in ein Hypertextsystem angeboten werden, sondern möglichst eng miteinander verzahnt werden" [Fuh90] sollen. Kuhlen [Kuh92] beschreibt verschiedene Arten von Browsing und Navigation, denen er die Verfahren des exact und partial match entgegenhält, wie sie in [BC87] ausführlich beschrieben sind, und hält die Symbiose von Hypertext und Information Retrieval für selbstverständlich.

Die Integration von externen Informationsquellen in Hypertextbasen bringt sowohl für den Umgang mit Hypertexten als auch für den Zugang zu Online-Diensten einen Mehrwert mit sich. Der Hypertext-Autor muß nicht jede, vielleicht selten benutzte Information in den Hypertext integrieren, sondern kann sie durch einen Verweis auf eine externe Quelle zugänglich machen. Der Nutzer von Online-Diensten kann mit Hilfe des Hypertextsystems Recherche-Ergebnisse archivieren, aber auch standardisierte Suchanfragen oder Eintrittspunkte in Datennetze nach seinem persönlichen Informationsbedarf im Hypertext organisieren [HR93, AHR93].

In anderen Systemen folgt die Integration externer Quellen im allgemeinen dem Browsing-Paradigma. Über sogenannte Query-Links wird externe Information durch Exploration erreicht [Kuh91]. Das KHS bietet einen der jeweiligen Informationsquelle angepaßten Interaktionsstil, indem Online-Datenbanken normalerweise durch Query-Formulierung, Gopher hingegen durch Exploration erschlossen werden.

Die Anpassung des Interaktionsstils erreicht das KHS durch eine Typisierung von Hypertext-Einheiten, sowie einen speziellen Tool-Kommunikationsmechanismus. Wir werden zunächst diese beiden Prinzipien in ihren Grundzügen einführen, bevor wir konkret auf Such- und Navigationsmöglichkeiten im KHS eingehen werden.

2 Das Hypertext-Modell des KHS

Das KHS ist als Vielzweck-Hypertextsystem konzipiert, das nicht auf eine spezielle Anwendung hin optimiert ist. Daher garantiert das KHS Uniformität der Präsentations- und Interaktionsstile, unabhängig von der jeweiligen Anwendung. Dies wird erreicht durch eine Menge von anwendungsunabhängigen Tools, deren Verhalten jeweils durch anwendungsspezifische Unit- bzw. Link-Typen parametrisiert wird. Durch eine Typisierung der Hypertext-Objekte [HK93], wie sie ähnlich auch in dem Hypertextsystem Concorde [HBS93] vorgesehen wird, wird darüber hinaus auch ermöglicht, anwendungsspezifische Konsistenzbedingungen in den Hypertext einzuführen.

Beispiele für die im KHS bereits realisierten Applikationen sind Thesaurusaufbau und -pflege (s. Abschnitt 6), die Verwaltung und Archivierung von e-mail, der Zugang zu Datenbankbeschreibungen, sowie die Archivierung der Ergebnisse von Online-Recherchen [HR93]. Online-Hilfe für Software-Systeme ist eine weitere Klasse von Applikationen, die am Beispiel eines "Hilfe-Hypertexts" für das KHS, zur Zeit im Rahmen einer Diplomarbeit erschlossen wird. Die Gegenstände der Applikation, z.B. Begriffe und Relationen eines Thesaurus, werden dabei auf spezielle Unit- und Link-Typen abgebildet (z.B. *Thesaurus-Item* als Unit-

Typ und *ISA* als Link-Typ). Konsistenzregeln beziehen sich in erster Linie auf Beziehungen zwischen derartigen Objekten, so darf ein ISA-Link z.B. nur Hypertext-Einheiten vom Typ *Thesaurus-Item* zyklensfrei verbinden.

2.1 Unit-Typen

Units repräsentieren die Inhalte des Hypertexts. Ihre Präsentations- und Interaktionsmöglichkeiten werden durch den jeweiligen Typ festgelegt. Gemeinsam ist allen Hypertext-Units, daß sie einen Namen haben und durch Verschlagwortung mit gewichteten Indextermen einer Indexsuche zugänglich gemacht werden können. Hinsichtlich der Typisierung lassen sich grundsätzlich die mediale Information enthaltenden terminalen und die Strukturinformation enthaltenden nonterminalen Units unterscheiden. Ein Sonderfall der Zuordnung von medialer Information ist, daß Ton — quasi als zusätzlicher Informationskanal — Units jeglichen Typs beigefügt werden kann.

Terminal Units regeln die Datenhaltung, Präsentation von und die Interaktion mit medialer Information. Exemplarisch sollen hier die wichtigsten Grundtypen terminaler Units aufgeführt werden:

1. **Text-Units** sind zur Darstellung konventionellen Fließtexts geeignet.
2. **Form-Units** erlauben den strukturierten Zugriff auf Information (analog einem Formular). Untertypen sind zur Führung von Adreßlisten oder bibliographischen Referenzen vorgesehen, oder zur Repräsentation von e-mail etc.
3. **Image-Units** bilden Rasterbilder ab.
4. **External-Application-Units** vermitteln den Zugang zu Applikationsprogrammen. Wird eine Unit dieses Typs während der Navigation erreicht, wird automatisch ein externes Programm, wie z.B. ein PostScript-Viewer oder ein Vektorgraphik-Editor, aktiviert und mit den erforderlichen Daten versorgt.

Composite Units sind ein von Links zu unterscheidendes Strukturierungsinstrument für Hypertexte, indem sie selbst weitere Hypertext-Units enthalten und dem Hypertext somit eine hierarchische, im Falle des KHS sogar polyhierarchische Grundstruktur aufprägen. Diese Struktur kann dann auf der Präsentationsebene z.B. zur Generierung von Inhaltsverzeichnissen genutzt werden. In dieser Hinsicht entsprechen Composite Units den Table-of-Content-Nodes von Textnet [TW86] oder den hierarchischen Dokumentstrukturen von Augment [Eng84]. Darüber hinaus kann durch Composite Units eine anwendungsspezifische Ordnung der Units (sequentiell nach Wahl des Autors, alphabetisch, nach Erzeugungszeit etc.) definiert werden, eine Eigenschaft, die mit Pfad-Mechanismen anderer Systeme verglichen werden kann. Ist eine Hypertext-Einheit in mehreren Composite Units enthalten, kann die Verfügbarkeit von speziellen Links während der Navigation davon abhängig gemacht werden, welcher Zweig der Polyhierarchie gerade durchwandert wird. Sowohl Suchoperationen als auch Link-Navigation können auf die durch eine oder mehrere Composite Units verbundenen Teilbäume der Strukturhierarchie beschränkt werden. In dieser Hinsicht sind Composite Units vergleichbar mit Kontexten, wenn sie auch nicht die Möglichkeit bieten, Revisions von Hypertext-Knoten vorzusehen [DS87].

2.2 Link-Typen

Während die Hypertext-Units mediale Information enthalten, repräsentieren Links Beziehungen zwischen Units, die durch den Inhalt oder auch den Verwendungszusammenhang begründet sein können. KHS-Links können grundsätzlich eine beliebige Stelligkeit haben und beliebige Hypertext-Objekte (Units, Links) verbinden. Im allgemeinen wird aber nur von binären gerichteten Links zwischen Units Gebrauch gemacht. Sowohl in der

Ausgangs- als auch der Zieleinheit können diese Links auf Teilinformation Bezug nehmen, die bei der Interaktion als *Hotwords* oder *Hotareas* (bei Graphik) zur Verfügung stehen.

Der Link-Typ spezifiziert formale Eigenschaften der durch Links diesen Typs repräsentierten Relation, wie z.B. Transitivität, Zyklentfreiheit, Symmetrie etc. Applikationsspezifische Konsistenzbedingungen können über Integritätsregeln definiert werden, die spezifizieren, unter welchen Bedingungen ein Link eines bestimmten Typs erzeugt werden darf. Links, die diese Bedingungen nicht mehr erfüllen, werden automatisch entfernt.

Da Link-Typen den Verwendungszusammenhang von Hypertext-Units repräsentieren, haben sie auch Einfluß auf die Navigationssemantik. Je nach Typ des traversierten Links wird z.B. entschieden, ob die Zielunit zusätzlich oder alternativ zur momentan präsentierten dargestellt wird.

3 Tools zur Implementation generischer Interaktionsformen

Das uniforme Erscheinungsbild von KHS-Applikationen wird durch ein baukastenartig kombinierbares System von Tools erreicht. All diese Tools, mit Ausnahme des KHS-Browsers (er zeigt die im Fokus liegende Unit, ihre Links und ihre Einbindung in die Polyhierarchie, s.a. Abschnitt 4.2), implementieren relativ elementare Präsentations- und Interaktionsformen, wie z.B.

1. Auswahl einer Unit aus einer Liste. Listeninhalte können Retrievalergebnisse, ungelesene e-mail oder Zielunits von Links sein.
2. Hierarchische Darstellungen von Hypertext-Strukturen, sei es die Strukturhierarchie oder die transitive Hülle von Links.
3. Präsentation des Inhalts einer Unit.

Jedes dieser Tools bietet also eine spezifische und damit eingeschränkte Sicht auf den Hypertext mit einer überschaubaren Anzahl von Funktionen. Eine hinlängliche Mächtigkeit erreicht KHS durch die Kooperation dieser Tools. In jeder Dialogsituation kann der Nutzer die Tools, die er zu benötigen meint, aktivieren. Jedes dieser Tools präsentiert dann seine Sicht auf ein und dasselbe Segment des Hypertexts. Etwaige Navigations- oder Änderungsoperationen führen dann zu angepaßten Änderungsoperationen der aktiven Tools. Tools, die in der erreichten Dialogsituation erwiesenermaßen überflüssig sind, z.B. weil die Liste der ungelesenen Mail während der Navigation völlig geleert wurde, werden automatisch entfernt.

4 Navigation

Als Navigation im Hypertext wollen wir im folgenden jede Interaktion verstehen, die auf einem schrittweisen, Link- und Strukturbeziehungen ausnutzenden Verfahren beruht. "Navigation" umfaßt also sowohl ein ungerichtetes "Stöbern" im Hypertext als auch zielgerichtetes sequentielles Verfolgen von Pfaden. In beiden Fällen wird die Navigation vorangetrieben, indem der Nutzer aus einem Angebot von Verknüpfungen eine Zielunit auswählt. Diese wird dann in einem vom System zu bestimmenden Kontext dargestellt. Im weiteren Verlauf dieses Abschnitts werden wir zunächst den Mechanismus zur Kontextauswahl kurz skizzieren und dann die wichtigsten Navigationstools beschreiben.

4.1 Wahl des Kontexts

Ist eine Hypertext-Unit in mehrere Kontexte eingebettet, das heißt, gibt es in der Polyhierarchie der Composite Units mehrere Pfade, die von der obersten Ebene zu dieser Unit führen, so stellt sich bei der Hypertext-Navigation die Frage nach der jeweils geeigneten

Auswahl des Kontexts, die aufgrund der jeweils unterschiedlichen verfügbaren Verknüpfungen erhebliche Auswirkungen auf die weiteren Navigationsmöglichkeiten haben kann.

KHS bietet zwei anwendungsunabhängige Verfahren der Kontextauswahl:

1. Der neue Kontext wird so gewählt, daß innerhalb der Strukturhierarchie ein möglichst geringer Abstand zum aktuellen Kontext besteht. Dieses Verfahren ermöglicht z.B. ein sequentielles Traversieren von Teilbäumen.
2. Der neue Kontext wird so gewählt, daß eine möglichst hohe Nähe zum Thema des bisherigen Hypertext-Dialogs gegeben ist. Das Thema des Dialogs wird zur Zeit aus einer evtl. gegebenen Suchanfrage ermittelt oder/und aus den Schlagworten, die den bisher traversierten Hypertext-Einheiten zugeordnet sind. Die inhaltliche Nähe ergibt sich somit allein aus der Kookkurrenz von Schlagworten, soll aber durch Einsatz eines Thesaurus (s. Abschnitt 6) differenzierter bestimmt werden. Ergibt sich für zwei Kontexte die gleiche Bewertung, so kann auf Verfahren 1 zurückgegriffen werden.

Die Dialoghistorie, die zur Ermittlung des Dialogthemas mitgeführt werden muß, kann auch explizit zur Navigation genutzt werden, indem vergangene Dialogzustände reaktiviert werden können.

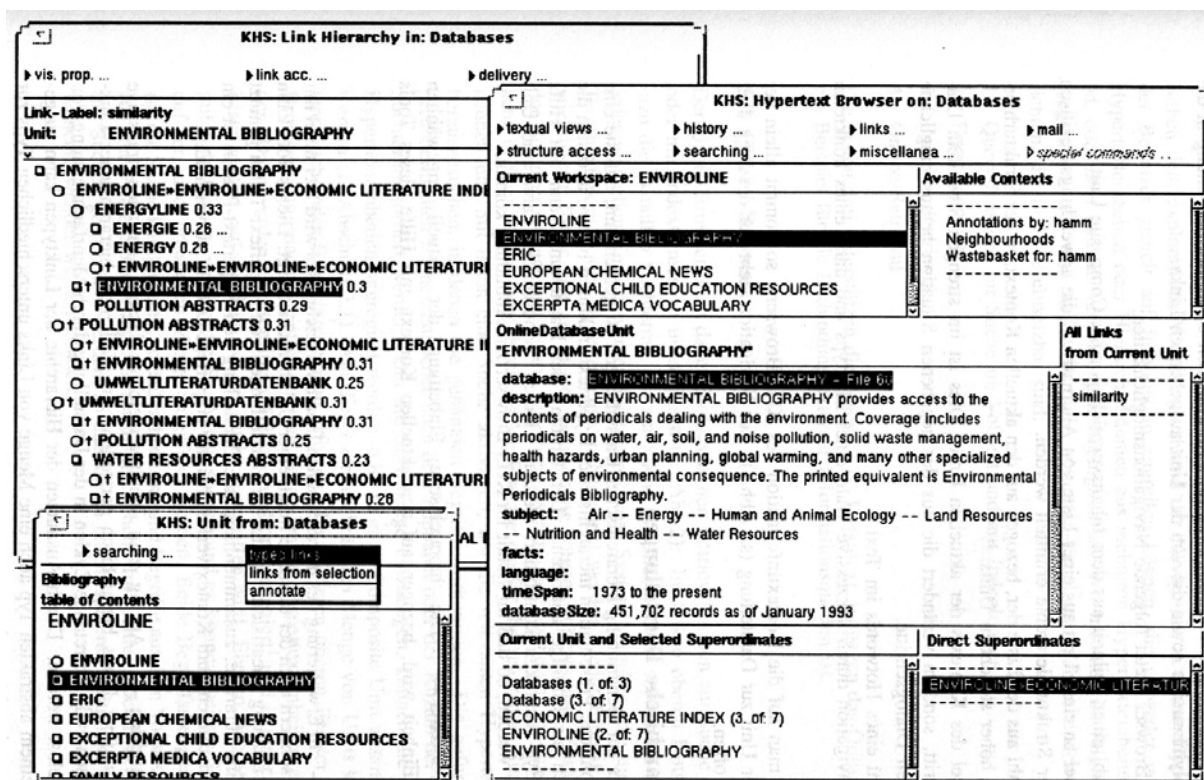


Abbildung 1 Im rechten Bildteil erkennt man den KHS-Hypertext-Browser mit der selektierten Unit „Environmental Bibliography“. Links sind zwei Browser zu sehen, die die Navigation des Benutzers durch die Darstellung der Link-Hierarchie bzw. der Hypertext-Hierarchie im augenblicklichen Kontext unterstützen sollen.

4.2 KHS-Browser

Der KHS-Browser (s.a. Abb. 1) nimmt insoweit eine Sonderstellung unter den KHS-Tools ein, als er als einziges Tool so ausgelegt ist, daß Hypertext-Navigations- und -Editieroperationen auch ohne Zuhilfenahme weiterer Tools durchgeführt werden können. Alle weiteren Tools werden von hier aus aktiviert. Der KHS-Browser präsentiert jeweils den Inhalt

der im aktuellen Dialogschritt erreichten Unit, ihre Einbettung in die Strukturpolyhierarchie sowie die von der Unit ausgehenden Links.

Der KHS-Browser bietet folgende Navigationsmöglichkeiten:

1. Auswahl einer Teilunit aus dem Inhaltsverzeichnis einer Composite Unit.
2. Wechsel zu einer Unit aus einer Liste von Alternativen, die anwendungsspezifisch aus der Strukturhierarchie ermittelt werden.
3. Auswahl aus der Liste der, bezogen auf den aktuellen Kontext, in der Strukturhierarchie höher stehenden Units.
4. Wechsel des Kontexts der aktuellen Unit. Dies ist im strengen Sinn kein Dialogschritt, sondern verändert die aus der gegebenen Situation heraus möglichen weiteren Dialogschritte.
5. Anwahl eines Hotwords im Text.
6. Linknavigation durch sukzessive Wahl eines Linktyps und dann eines konkreten Links.

Verzichtet man auf die Kontextinformation des KHS-Browsers, so kommt allein der Inhalt einer Unit zur Darstellung (s. Abb.1). Zur Navigation steht nur der Weg über Hotwords offen.

4.3 Hierarchische Darstellungen

Allen Tools zur Generierung hierarchischer Darstellungen ist gemeinsam, daß sie eine pseudographische Präsentation hierarchisierter Hypertextstrukturen bieten, in denen die Unit-Namen als Einträge auftreten. Der Umfang der Darstellung kann mit Hilfe von Fish-Eye-Techniken kontrolliert werden. Durch "Mouseclick" kann eine Unit ausgewählt werden, so daß ihr Inhalt im KHS-Browser zur Abbildung kommt.

Table of Contents — Die hierarchische Einbettung der aktuell angewählten Hypertext-Einheit wird, bezogen auf den aktuellen Kontext, mit Hilfe dieses Tools dargestellt.

Volltext — In Erweiterung des hierarchischen Inhaltsverzeichnisses wird hier nicht nur der Name, sondern auch der textuelle Inhalt einer Unit berücksichtigt. Dieses Tool kann also zur Erzeugung linearisierter textueller Darstellungen von Hypertext-Teilen genutzt werden. Eine derartige, zusammenhängende Darstellung von Hypertext-Pfaden ist ein wichtiges Instrument, um Kontextverlust bei der Hypertext-Lektüre zu vermeiden.

Link-Hierarchien — Analog zum Inhaltsverzeichnis lassen sich auch transitive Beziehungen über Links hierarchisch darstellen. Im Fall nicht-hierarchischer Link-Typen müssen evtl. auftretende Zyklen in der Darstellung durch Mehrfachnennung von Units aufgelöst werden. Durch Ausnutzen der Hierarchie der Linktypen kann ausgehend von einem abstrakten Typ auch eine Menge von Links unterschiedlichen Typs zur Darstellung gebracht werden.

5 Such- und Retrievalinstrumente im KHS

Neben dem explorativen Vorgehen während der Navigation in einem Hypertext muß der Benutzer gezielt auf Information zugreifen können, die entweder in einzelnen Hypertexteinheiten oder Hierarchieebenen aus mehreren Hypertexteinheiten abgelegt ist, oder über externe Verbindungen zugänglich ist. Wir möchten im folgenden unser Augenmerk auf vier Arten des dem Matching-Paradigma gehorchenden Retrievals richten, die im KHS einsetzbar sind:

1. Query-Browser zur Suche im Volltext und mit Indextermen

2. Formularsuche
3. Passagen-Retrieval
4. Online-Retrieval

Die ersten drei unterstützen die Auswahl im Hypertext, wogegen das Online-Retrieval ein Beispiel für die Einbindung externer Informationsquellen ist.

5.1 Query-Browser

Der Query-Browser dient dazu, die klassischen Retrievalformen zur Suche im Volltext oder mit Indextermen zu unterstützen (s. Abb.2). Im linken oberen Fenster erkennt man die aktuellen Suchterme, die zur Suche mit Indextermen verwendet werden. Zu jeder textuellen Einheit werden Indexterme vom System vergeben, die aufgrund von Termhäufigkeit in der Hypertexteinheit, berechnet werden. Jeder Einheit werden die wichtigsten Terme, die aus dieser Berechnung folgen, zugewiesen. Zusätzlich kann der Benutzer weitere Indexterme vergeben, löschen oder verändern. Im Query-Browser kann er über Menüs oder durch direkte Eingabe die für ihn relevanten Suchterme benennen und mit ihnen eine Suche aktivieren, um die relevanten Hypertexteinheiten herauszusuchen, in denen die genannten Begriffe vorkommen. Die gefundenen Hypertexteinheiten können entweder nach dem Alphabet oder der Anzahl der gefundenen Suchterme sortiert werden. Um die Suchmenge einzuengen, können die gefundenen Hypertexteinheiten aggregiert werden, d.h. in einer Composite Unit zusammengefaßt werden (s. Abschnitt 2.1). Auf dieser eingeschränkten Menge von Units können weitere Such- oder Browsingaktionen erfolgen. Im mittleren oberen Fenster des Query-Browsers kann nach Begriffen aus den Textteilen der Hypertexteinheiten gesucht werden. Dazu werden alle Begriffe auf ihre Grundform reduziert² und in einer invertierten Liste geführt, auf die dann bei der Suche zugegriffen wird. Ebenso, wie bei der Suche mit Indextermen, werden die Ergebnisse sortiert und es können weitere Suchvorgänge auf der zuvor erhaltenen Menge ausgeführt werden. Beide Begriffsmengen lassen sich für die Suche kombinieren, so daß Indexterme und Begriffe aus dem Volltext gemeinsam für eine Frageformulierung zur Verfügung stehen. Im rechten oberen Fenster können die Suchen gespeichert werden, so daß sie zu einem späteren Zeitpunkt wieder- oder weiterverwendet werden können.

5.2 Formularsuche

Zu den im KHS benutzbaren Strukturierungsalternativen gehören die in Abschnitt 2.1 genannten Form-Units. Mit KHS kann in Feldern solcher strukturierten Hypertexteinheiten gezielt gesucht werden. Dazu wird ein Form-Search-Browser geöffnet, in dem der Typ der Einheiten, in denen gesucht werden soll, bestimmt wird, und die gesuchten Begriffe in die entsprechenden Felder eingegeben werden (beispielsweise einen Autorennamen in das Autorenfeld einer Publikationseinheit oder die Telefonnummer in ein Telefonfeld einer Adresseneinheit). Der Nutzer ist bei der Wahl der Hypertexteinheiten nicht auf einen einzelnen Typ von Hypertexteinheiten beschränkt, sondern kann auch Oberklassen von Typen zur Suche verwenden. Er kann etwa nach Autoren einer Publikation in sogenannten Articleinheiten suchen, aber auch im Bereich Publikationseinheiten, der alle Arten von Publikationen, also auch Buch- und Artikelbeschreibungen, enthält. Die Suche erfolgt dann in den spezifizierten Unit-Typen. Die Ausgabe der gefundenen Einheiten erfolgt, wie im Query-Browser, als eine Liste der relevanten Units, die durch Anklicken im KHS-Browser betrachtet werden können.

² Dazu wurde ein Verfahren nach Kuhlen [Kuh77] verwendet.

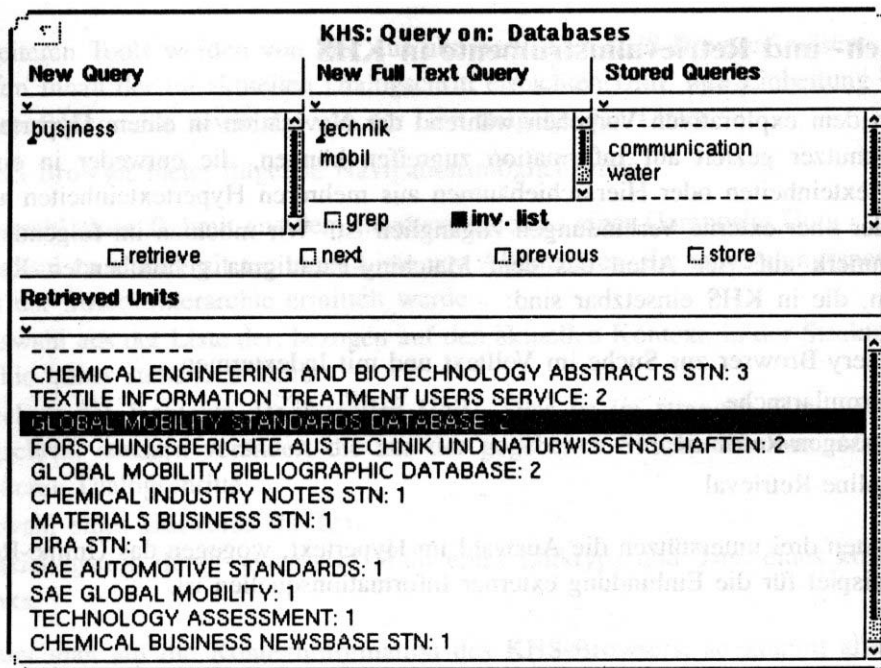


Abbildung 2 Der Query-Browser mit den Fenstern zur Suche mit Indextermen, zur Volltextsuche, zum Abspeichern von Suchfragen und zum Anzeigen der aktuellen Suchergebnisse.

5.3 Passagenretrieval

Im KHS werden alle Einheiten eines Hypertextes mit Hilfe von Ähnlichkeitsmaßen miteinander verknüpft. Diese Verknüpfungen basieren auf statistischen Berechnungen, die mit auf Stammformen reduzierten Termen arbeiten. Dabei werden die inverse Dokumenthäufigkeit und das Cosinus-Maß [SM87] zur Identifikation ähnlicher Texte herangezogen. Um aber Ähnlichkeiten beliebiger Textpassagen zu finden, kann in einer Hypertexteinheit ein Textteil beliebiger Länge markiert werden, welcher dann als eine Fragestellung verwendet wird. Die Terme aus dem selektierten Textteil werden auf ihre Stammform reduziert, die Stopwörter entfernt und zu dem Textteil ähnliche Hypertext-Einheiten gesucht. Dabei kann entweder auf das Cosinus-Maß oder den asymmetrischen Koeffizienten zur Berechnung der Ähnlichkeit zurückgegriffen werden, die beide als geeignet für die Zwecke des Passagen-Retrieval angesehen werden können (siehe [KH93]).

5.4 Online-Retrieval

Genügen die augenblicklich im Hypertext verfügbaren Ressourcen nicht mehr, um das Informationsbedürfnis zu befriedigen, so kann der Benutzer auf externe Quellen, in diesem Fall Online-Datenbanken, zurückgreifen³. Durch den Einsatz des Online-Database-Browsers versucht KHS das Informationsproblem zu lösen. Der Benutzer kann den Online-Database-Browser an jeder beliebigen Stelle im KHS aktivieren. KHS generiert mit vom Benutzer ausgewählten Suchtermen eine auf Boole'scher Logik basierende Fragestellung. Diese wird auf einer externen Datenbank bearbeitet und die Ergebnisse an KHS übergeben. Dort werden sie in die polyhierarchische Hypertextstruktur integriert und dem Benutzer zur Relevanzentscheidung präsentiert.

³ Dabei wurde ein Verfahren entwickelt, wie es in [Smi88, PM89] gefordert wurde und in [Kuh91, Rit93] Lösungsansätze beschrieben werden. Ausführlich diskutiert wird die Realisierung mit dem KHS in [HR93, AHR93]. Neben der Einbindung von Online-Recherchen im KHS wird dort auch ausführlich auf die Integration weiterer Mehrwertdienste, wie e-mail und Gopher eingegangen. Über Gopher können natürlich wieder weitere Mehrwertdienste, beispielsweise OPACs oder WAIS-Datenbanken angesprochen werden [Aßf93].

6 Einsatz von Thesauri in Suche und Navigation

In einem Hypertextsystem wird ein Thesaurus primär bei der Verarbeitung einzelner Terme eingesetzt. Im KHS ist dies vor allem bei der retrievalgestützten Interaktion und teilweise auch bei der Navigation der Fall. Durch die in Abschnitt 6.2 besprochene Vorgehensweise kann der Thesaurus jedoch auch zur Verarbeitung ganzer Texteinheiten eingesetzt werden.

6.1 Thesauri

Thesauri enthalten ein kontrolliertes Vokabular und die zwischen den (das Vokabular bildenden) Begriffen bestehenden Relationen. Wichtige Relationen sind z.B. Hierarchie-Relationen wie *Ober-/Unterbegriff (OA/UA)* oder *Verbands-/Teilbegriff (SP/TP)* die je nach Thesaurus eine strikte Hierarchie oder eine Polyhierarchie (wie im KHS) aufspannen, und somit die Begriffe ordnen. Hauptanwendungsgebiet von Thesauri ist die Indexierung und Suche.

Im KHS existiert für den Thesaurus ein bestimmter Unit-Typ. Alle Einträge des Thesaurus (die Begriffe) sind Units von diesem Unit-Typ. Der Thesaurus ist somit ein Teil des Hypertextes, in dem wie gewohnt navigiert und editiert werden kann. Die Relationen werden durch Links realisiert. Hierzu stellt das KHS für jede Relation ein eigenen Link-Typ bereit. Die in den Abschnitten 2.1 und 2.2 besprochenen Konsistenzbedingungen legen fest, wo die Thesaurus-Units auftreten dürfen, und garantieren, daß Thesaurus-Links (die Relationen) nur zwischen Thesaurus-Units auftreten.

Eine weitere Besonderheit des Thesaurus im KHS ist, daß jeder Thesaurus-Eintrag im KHS eine *scope note* besitzt. Diese Scope-Notes werden wie bei einem Glossar zur Erläuterung benutzt (s.a. Abschnitt 6.3). Somit ist im KHS das Glossar im Thesaurus integriert.

6.2 Thesauri und Suche

Mit Hilfe eines Thesaurus kann die Suche auf vielerlei Art und Weise verbessert werden. Ansatzpunkt im KHS ist dabei immer, daß die Suche anhand von Termen (bzw. Indextermen) durchgeführt wird. Diese Terme können unterschiedlicher Herkunft sein. So werden sie bei einer normalen Suchanfrage vom Benutzer eingegeben, bei generierten Suchanfragen werden es Indexterme oder aus einer Texteinheit abgeleitete Terme sein.

Aufgrund des großen Wortschatzes, den die natürliche Sprache bietet, ist es unwahrscheinlich, daß von allen an einem Hypertext beteiligten Personen und Programmen das gleiche Vokabular verwendet wird. Da das Retrieval im KHS (und nicht nur da) auf dem Matching-Paradigma basiert, kann dies dazu führen, daß relevante Texteinheiten nicht gefunden werden, da aufgrund des unterschiedlichen Vokabulars das Matching fehlschlägt. Hier hilft ein Thesaurus weiter. Er legt ein kontrolliertes Vokabular fest, das alle Beteiligten verwenden müssen.

Durch die Verwendung der im Thesaurus vorhandenen Relationen kann eine Suchfrageerweiterung durchgeführt werden. Die Relation *Benutze Synonym/Benutzt für (BS/BF)* erlaubt das Erkennen bedeutungsgleicher Terme⁴, die weiteren Thesaurus-Relationen (*Ober-/Unterbegriff(OA/UA)*, *Verwandter Begriff (VB)*, usw.) erlauben den Abgleich von semantisch verwandten Termen. So ist es möglich, Texteinheiten zu finden, die ein verwandtes Thema behandeln. Herkömmliche Suchfrageerweiterungen nutzen die hierarchischen Relationen exzessiv aus, d.h. es werden alle Stufen einer hierarchischen Relationen betrachtet. Dies erhöht den Recall beträchtlich, nicht unbedingt aber die Precision. Sinnvoll ist eine

⁴ Dies erlaubt in beschränktem Maße doch Varianten im Wortgebrauch, was u.a. die Akzeptanz der Software durch den Benutzer erhöht.

Einschränkung der Erweiterung, bei der nur die näheren Terme bzgl. der Hierarchierelationen mit einbezogen werden [CK87]. Durch die im folgenden beschriebene Vorgehensweise ist es sogar möglich Aussagen über die Beziehung zwischen Texteinheiten zu machen, so daß der Benutzer als Ergebnis einer Suche nicht nur Texteinheiten präsentiert bekommt, die ein verwandtes Thema behandeln, sondern auch die Relation in welcher das Thema der gefundenen Texteinheit und das aktuelle Thema zueinander stehen. Somit wird nicht nur der Recall der Suche erhöht (aber nicht so stark wie bei herkömmlichen Suchfrageerweiterungen), sondern auch die Precision (allerdings nicht im globalen Sinne, da viele gefundene Einheiten vom Suchthema abweichen ohne zu interessieren; aber der Benutzer findet leichter die für ihn relevanten Ergebnisse).

Doch nun zu der Aufgabe, Beziehungen zwischen Texteinheiten unter Zuhilfenahme des Thesaurus herauszufinden (die in diesen und den nächsten Abschnitt immer wiederkehrt). Hierzu müssen für die Texteinheiten repräsentative Mengen von Termen gefunden werden (normalerweise sind dies die Indexterme; beim Passagenretrieval (s. Abschnitt 5.3) werden die Terme allerdings direkt aus dem Text generiert), die das Thema der Texteinheiten darstellen. Dann wird versucht, aus den Relationen zwischen einzelnen Termen auf Relationen zwischen den Themen zu schließen. Wenn z.B. *Thema-1* fast nur Terme enthält, die den Termen von *Thema-2* bezüglich der Relation *Ober-/Unterbegriff(OA/UA)* übergeordnet sind, kann man schließen, daß *Thema-1* *Thema-2* übergeordnet ist. Die Schwierigkeit hierbei ist der Übergang (die Schlußfolgerung) von Relationen zwischen einzelnen Termen auf Beziehung zwischen Mengen von Termen. Dies kann nur heuristisch geschehen. Hierzu wird im KHS die im Aufbau befindliche Wissensbasis herangezogen werden.

Diese Vorgehensweise zur Generierung von Beziehungen zwischen Mengen von Termen ist auch zur automatischen Erzeugung von weiterleitenden (z.B. vertiefenden) Suchanfragen einsetzbar. Klassifiziert der Benutzer z.B. das Ergebnis der Suche, kann versucht werden, aus den positiv klassifizierten Texteinheiten ein gemeinsames Thema herauszufinden, das in eine neue (automatisch generierte) Suchanfrage mündet (Relevanz-Feedback). Hierbei sind die Möglichkeiten nur durch die in dem benützten Thesaurus verfügbaren Terme und Relationen eingeschränkt.

6.3 Thesauri und Navigation

In der Navigation findet der Thesaurus vor allem Anwendung im Bereich der Linkhandhabung. Ein einfaches Beispiel ist das Glossar im KHS. Im KHS werden automatisch Links zwischen Termen im aktuellen Text und erklärenden Scope-Notes im Thesaurus erzeugt (*Glossary-Links*). Da die Scope-Notes im Thesaurus stehen, müssen hierzu nur Links zwischen den Termen der aktuellen Texteinheit und den entsprechenden Thesaurus-Einträgen aufgebaut werden, denen der Benutzer bei Bedarf folgen kann. Die *Glossary-Links* sind im KHS nur temporär, d.h. der Aufbau der *Glossary-Links* erfolgt bei der Navigation zu einer Textunit. Navigiert der Benutzer weiter, so verschwinden diese *Glossary-Links* wieder, und neue werden erzeugt.

Das in Abschnitt 5.3 besprochene Passagenretrieval wird im KHS zur automatischen Erzeugung von Links eingesetzt. Da das Passagenretrieval auch auf dem Matching-Paradigma beruht, kann es durch Einsatz des Thesaurus differenzierter gestaltet werden, was zu einer differenzierteren Linkerzeugung führt.

Ein weiteres Einsatzgebiet ist das Anbieten von Zusatzinformationen über Links. In vielen Hypertextsystemen beschränkt sich die Navigation auf die Wahl eines der möglichen Links, ohne daß der Benutzer Informationen über das Ziel des Links bekommt. Die von Landow [Lan87] gewählte Bezeichnung für dieses Themengebiet ist *Rhetoric of Arrival*. Thesauri

können dazu benutzt werden, Zusatzinformationen zu Links anzubieten. Beispielsweise kann man zu jedem Link der aktuellen Unit die Indexterme der Zielunits bestimmen, und dann nach dem Verfahren des letzten Abschnitts die Beziehung zwischen der aktuellen Unit und der Zielunit ermitteln und dem Benutzer präsentieren.

Wenn man diese Zusatzinformation über Links mit den Informationen verknüpft, die man durch die Unit-Typen, die Link-Typen und der Dialog-Historie von KHS bekommt, ist es möglich Links nach dem Grade des Interesses für den aktuellen Benutzer zu sortieren und zu filtern. Eine einfache Möglichkeit wäre z.B. nur noch Links zu thematisch verwandten Units anzuzeigen.

6.4 Weitere Einsatzgebiete von Thesauri in Hypertextsystemen

Ein weiteres Anwendungsgebiet für Thesauri im KHS ist das Klassifizieren von Units. Diese Klassifizierung kann durch den Thesaurus unterstützt bzw. verbessert werden, da es derzeit auf einem einfachen Matching beruht. Die Klassifizierung wird benutzt, um Units bestimmten Composite Units zuzuteilen. Eine Anwendung hierfür ist das Einordnen von einkommender e-mail. Diese wird im KHS anhand der im Titelfeld vorkommenden Terme klassifiziert und eingeordnet, so daß der Benutzer seine e-mail schon vorsortiert bekommt.

7 Ausblick

Die obigen Ausführungen geben den aktuellen Stand der Arbeiten am KHS wieder. Abschließend soll nun ein kurzer Ausblick auf die künftige Entwicklung des Systems hinsichtlich der hier angesprochenen Fragestellungen gegeben werden. Das Retrieval-Modell für Hypertexte wird einerseits in Richtung auf eine weitere Vereinheitlichung der Interaktion weiterentwickelt, andererseits durch eine vermehrte Ausnutzung von Hypertext-Strukturen für die Suche im Sinne eines kaskadierten Retrieval in Anlehnung an die im Rahmen des Projekts TOPOGRAPHIC [KHST89] entwickelten Verfahren verbessert. Eine verbesserte Integration externer Informationsquellen basiert sowohl auf einem vereinheitlichten Interaktionskonzept als auch auf einer wissensbasierten Auswahl der relevanten Quellen. Dementsprechend wird KHS um eine Wissensbasis erweitert, in die der Thesaurus, wie schon von Defude ausgeführt [Def84], auch eingebettet wird.

Literaturverzeichnis

- [Ago88] M. Agosti: Is hypertext a new model of information retrieval? In: Online Information 88. Proceedings, 12. International Online Information meeting. Vol. 7, Seite 57-62, 1988.
- [AHR93] R. Aßfalg, R. Hammwöhner und M. Rittberger: The Hypertext Internet Connection: E-mail, Online Search, Gopher. In: D. Raitt and B. Jeapes (eds.), Online Information 93. 17th International Online Information Meeting, 7-9. December, London, pp. 453-464. Learned Information Ltd, 1993.
- [Aßf93] R. Aßfalg: Aspekte der Integration von Internetdiensten. In diesem Band, Seite 68-80, 1993.
- [BC87] N.J. Belkin und W.B. Croft: Retrieval Techniques. Annual Review of Information Science and Technology (ARIST), Band 22, Seite 110-145, 1987.
- [CK87] P.R. Cohen und R. Kjeldsen: Information Retrieval by Constrained Spreading Activation in Semantic Networks. Information Processing & Management, Vol. 23(4), Seite 255-268, 1987.

- [Def84] B. Defude: Knowledge based systems versus thesaurus: An architecture problem about expert systems design. In: CJ. van Rijsbergen, Hrsg., Research and development in information retrieval. Proceedings of the 3rd joined BCS and ACM Symposium., Seite 267 - 280, 1984.
- [DS87] N. Delisle und M. Schwartz: Contexts: a Partitioning Concept for Hypertexts. In: Computer Supported Cooperative Work Conference, 1987.
- [Eng84] D.C. Englebart: Authorship provisions in Augment. In: Proceedings of the IEEE COMPCON, Seite 465-472, 1984.
- [Fuh90] N. Fuhr: *Hypertext und Information Retrieval*. In: P.A. Gloor und N.A. Streitz, Hrsg., *Hypertext und Hypermedia. Von theoretischen Konzepten zur praktischen Anwendung*, Aus der Serie: Informatik Fachberichte 249, Seite 101-111. Berlin: Springer, 1990.
- [HBS93] M. Hofmann, E. Bartscht und S. Schmezko: *Klassendefinition und Anfrageformulierung in Informationssystemen durch graphische Interaktion*. In: G. Knorz, J. Krause und C. Womser-Hacker, Hrsg., *Information Retrieval'93. Von der Modellierung zur Anwendung*, Heft 12 in Schriften zur Informationswissenschaft, Seite 242-254. Universitätsverlag Konstanz, 1993. Proceedings der I. Tagung Information Retrieval'93.
- [HK93] R. Hammwöhner und R. Kuhlen: *Semantic control to open hypertext Systems by typed objects*. Technical Report 34-93 (WITH-6/93), Universität Konstanz, Informationswissenschaft, 1993.
- [HR93] R. Hammwöhner und M. Rittberger: *KHS - ein offenes Hypertext-System*. In: G. Knorz, J. Krause und C. Womser-Hacker, Hrsg., *Information Retrieval'93. Von der Modellierung zur Anwendung*, Heft 12 in Schriften zur Informationswissenschaft, Seite 208 - 222. Universitätsverlag Konstanz, 1993. Proceedings der I. Tagung Information Retrieval'93.
- [KH93] R. Kuhlen und M.S. Hess: *Passagen-Retrieval - auch eine Möglichkeit der automatischen Verknüpfung in Hypertexten*. In: G. Knorz, J. Krause und C. Womser-Hacker, Hrsg., *Information Retrieval'93. Von der Modellierung zur Anwendung*, Heft 12 in Schriften zur Informationswissenschaft, Seite 100 - 115. Universitätsverlag Konstanz, 1993. Proceedings der I. Tagung Information Retrieval'93.
- [KHST89] R. Kuhlen, R. Hammwöhner, G. Sonnenberger und U. Thiel: *TWRM-TOPOGRAPHIC: Ein wissensbasiertes System zur situationsgerechten Aufbereitung und Präsentation von Textinformation in graphischen Retrievaldialogen*. Informatik Forschung und Entwicklung, Vol. 4(2), Seite 89-107, 1989.
- [Kuh77] R. Kuhlen: *Experimentelle Morphologie in der Informationswissenschaft*. Aus der Serie: DGD-Schriftenreihe Band 5. München: Dokumentation, 1977.
- [Kuh91] R. Kuhlen: *Aktivierung von Online-Informationsbanken aus Hypertextbasen*. In: S. Sorg, Hrsg., *Online '91. 14. Europäische Congressmesse für Technische Kommunikation. Congress IV. Bürokommunikation: Konzepte und Strategien zur Unterstützung der Büroarbeit*, Seite IV.01.01-IV.01.14, 1991.
- [Kuh92] R. Kuhlen: *Hypertext und Information Retrieval - mehr als Browsing und Suche*. In: R. Kuhlen, Hrsg., *Experimentelles und praktisches Information Retrieval*, Heft 3 in Schriften zur Informationswissenschaft, Seite 309-324. Universitätsverlag Konstanz: Konstanz, 1992.

- [Lan87] G.P. Landow: *Relationally Encoded Links and the Rhetoric of Hypertext*. In: *Hypertext '87 Papers*, Chapel Hill, NC, University of North Carolina, Seite 331-344, 1987.
- [PM89] M. Percival und N. Mac Morrow: *Evaluating the feasibility of using hypercard as an interface prototyping tool with reference to online Services: the impact of LSDN*. In: *Online Information 89. 13th International Online Information Meeting Proceedings*, Seite 265-276. Oxford: Learned Information, 1989.
- [RU93] M. Rittberger: *Online Retrieval und Hypertext: Auf dem Weg zu verknüpften Datenbanken und offenen Hypertextsystemen*. Technical report, 24-93 [WITH-2/93], Informationswissenschaft, Universität Konstanz, 1993.
- [SM87] G. Salton und M.J. McGill: *Information Retrieval - Grundlegendes für Informationswissenschaftler*. New York: McGraw-Hill, 1987.
- [Smi88] K.B. Smith: Hypertext - Linking to the future. *Online*, Vol. 12(3), Seite 32-40, 1988.
- [TW86] R.H. Trigg und M. Weiser: TEXTNET: A Network-Based Approach to Text Handling. *ACM Transactions on Office Information Systems*, Vol. 4(1), Seite 1-23, 1986.