

# Hypertext und Information Retrieval - mehr als Browsing und Suche

Rainer Kuhlen

Informationswissenschaft

Universität Konstanz

Postfach 5560 - D-7750 Konstanz

email:kuhlen@lindau.inf-wiss.fyp.uni-konstanz.de

## Zusammenfassung

Zwischen Hypertext und Information Retrieval wurden über die beiden Paradigmen des Browsing und des Matching sich ausschließende oder - in neuerer Zeit - auch sich ergänzende Unterschiede aufgebaut. Informationelle Mehrwerte durch die Kombination oder die Integration beider Informationssystemansätze können aber nicht nur durch die unterschiedlichen Navigationsformen erzielt werden, sondern auch durch die Übertragung des breiten jeweiligen Methodenspektrums. In diesem Beitrag wird dies an zwei Beispielen verdeutlicht: zum einen durch die Einbindung der reichen informationswissenschaftlichen Abstracting-Tradition in eine "rhetoric of arrival" bei Hypertext und zum andern durch eine Erweiterung von Hypertext in Richtung offener Informationssysteme, hier durch einen Vorschlag zur Realisierung von Frageverknüpfungen ("query links") bzw. zur Einbettung externer Textbestände in eine bestehende Hypertextbasis. Dazu werden knapp das Systemdesign und die Systemumgebung von WITH, ein in Konstanz in Entwicklung befindliches webbasiertes, textorientiertes Hypertextsystem, vorgestellt.

## Inhalt

1. Hypertext und Information Retrieval
2. Browsing- und Matching-Paradigma
3. Informationelle Mehrwerte
- 3.1 Mehrwerteffekte durch die Verwendung von Zusammenfassungen bei Hypertext
- 3.2 Verbindung von Hypertext zu On-line-Informationsbanken
- 3.3 Systemdesign von WITH

aus: Festschrift für Gerhard Lustyj  
Experimentelles und praktisches Informationelles  
Retrieval, hg. R. Kuhlen  
Universitätsverlag Konstanz: Konstanz 1992

- Lustig, G. (Ed.) [1986]. *Automatische Indexierung zwischen Forschung und Anwendung*. Hildesheim etc.: G. Olms (Linguistische Datenverarbeitung 5).
- Nagao, M. [1988]. Panel: Language Engineering - The Real Bottle Neck of Natural Language Processing. *COLING Budapest - Proceedings of the 12th International Conference on Computational Linguistics (COLING '88)*. Vol. 2. Budapest, 22-27 August, 1988. Budapest: John von Neumann Society for Computing Sciences, pp.448-453.
- Reimer, U. & U. Hahn [1989]. Das Modell der Textkondensierung im TOPIC-System: Textgraphen als methodische Grundlage für die variable Verdichtung und inhaltliche Ansteuerung von Textwissen in Volltextdatenbanken. In W. Lenders (Ed.), *Linguistische Datenverarbeitung und Neue Medien*. Tübingen: G. Narr, pp.45-68 (forum Angewandte Linguistik 17).
- Rostek, L. [1979]. Methoden des partiellen Parsing für das automatische Indexing: Syntaxgraphen zur Analyse von Sprachmustern. In R. Kuhlen (Ed.), *Datenbasen - Datenbanken - Netzwerke. Praxis des Information Retrieval. Vol. 1: Aufbau von Datenbasen*. München etc.: Saur, pp.251-282.
- Sager, N. [1978]. Natural Language Information Formatting: The Automatic Conversion of Text to a Structured Data Base. In M.C. Yovits (Ed.), *Advances in Computers. Vol. 17*. 1978. New York etc.: Academic Pr., pp.89-162.
- Sager, N. [1981]. *Natural Language Information Processing: A Computer Grammar of English and Its Applications*. Reading, Mass.: Addison-Wesley.
- Salton, G. (Ed.) [1971]. *The SMART Retrieval System: Experiments in Automatic Document Processing*. Englewood Cliffs/NJ: Prentice-Hall.
- Stevens, M.E.; V.E. Giuliano & L.B. Heilprin (Eds.) [1965]. *Statistical Association Methods for Mechanized Documentation. Symposium Proceedings*. Washington 1964. Washington, D.C.: U.S. Government Printing Office (National Bureau of Standards Miscellaneous Publication 269).
- Wahlster, W. [1986a]. The Role of Natural Language in Advanced Knowledge-Based Systems. In H. Winter (Ed.), *Artificial Intelligence and Man-Machine Systems. Proceedings of an International Seminar Organized by Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt (DFVLR)*. Bonn, Germany, May, 1986. Berlin etc.: Springer, pp.62-83 (Lecture Notes in Control and Information Sciences 80).
- Wahlster, W. [1986b]. Panel: Natural Language Interfaces - Ready for Commercial Success? *COLING '86 - Proceedings of the 11th International Conference on Computational Linguistics*. Bonn, August 25-29, 1986. Bonn: Institut für angewandte Kommunikations- und Sprachforschung (IKS), pp.161-167.
- Weischedel, R.M. & N.K. Sondheimer [1983]. Meta-Rules as a Basis for Processing Ill-Formed Input. *American Journal of Computational Linguistics*, 9 (3-4), pp.161-177.
- Woods, W.A. [1977]. Lunar Rocks in Natural English: Explorations in Natural Language Question Answering. In A. Zampolli (Ed.), *Linguistic Structures Processing*. Amsterdam etc.: North-Holland, pp.521-569.

## 1. Hypertext und Information Retrieval

Informationssysteme - was sonst sollten diese beiden Ausprägungen, Hypertext und Information Retrieval, sein? Natürlich kann das Navigieren in Hypertext auch nur "fun" sein: "Storyspace" und "Afternoon"; zwei Muster des Genre "Hyperfiktion", werden sicher nicht aus Informationsgründen durchstöbert. Der "normale" Ausgangspunkt, ein Retrieval-System zu befragen oder in ein Hypertextsystem einzusteigen, wird jedoch ein Informationsproblem sein. Für welches Informationsproblem ist Hypertext geeignet, für welches ein Information-Retrieval-System? Oder wird Hypertext erst dann überhaupt zu einem Informationssystem mit ausreichendem informationellen Mehrwert für einen Benutzer, wenn Hypertextsysteme Anleihen beim Information Retrieval machen, sei es nun über die Verwendung von Techniken der Inhaltserschließung oder von gezielteren oder inferierenden Suchtechniken oder durch die dem (On-line-)Information-Retrieval inhärente Offenheit? Und gilt das gleiche für klassische Retrieval-Systeme, wenn sie durch entsprechende Hypertext-Techniken, sprich freies Navigieren über assoziative oder semantisch definierte (typisierte) Verknüpfungen oder durch entsprechende graphische Interaktionsmöglichkeiten oder durch informierende Metainformationen bzw. Orientierungsmöglichkeiten angereichert werden?

Wir vermuten also über das Konzept des informationellen Mehrwertes einige Überlappungen zwischen beiden Gebieten, die über die bloßen Gegensätze von Browsing und Suche hinausgehen. Eine auf die Jahre 1990 und 1991 (bis 9/91) beschränkte Online-Recherche in einschlägigen Datenbanken zum Thema: "(Hypertext OR Hypermedia) AND (Information ADJ Retrieval)" bestätigte die Vermutung und erbrachte zunächst einmal die folgenden quantitativen Resultate:

Compuscience	31
INFODATA	45
INSPEC	120
Computer Database	6
NTIS	13
Microcomp.Index	18
Eric	9
LISA	43
ISA	25

Insgesamt, unter Berücksichtigung von Mehrfachnennungen, ca. 200 Artikel aus den beiden Jahren 1990 und 1991 zum engeren Überlappungsbereich. Ist man allgemein an Literatur zu Hypertext und Hypermedia interessiert, so wird man schon ca. 2500 Fundstellen durchsuchen müssen. Allein im Zeitraum 1990, 1991 bei INSPEC 718, insgesamt bei INSPEC schon 1187 Treffer. Viele, aber doch noch wenige im Verhältnis zu den Dokumenten, die sich allein bei INSPEC mit Information Retrieval beschäftigen, nämlich seit Januar 1969 8328.

Keine Frage also, daß das Thema Hypertext und Information Retrieval die Fachwelt beschäftigt hat<sup>1</sup>. Was alles wird damit angesprochen: die Kombination unterschiedlicher Informationstypen; die Konversion von Text in Hypertext durch Inhaltsschließungstechniken; die Verwaltung von On-line-Thesauri durch Hypertext, auch die Verwaltung von (framebasierten) Wissensbeständen; Möglichkeiten des Absatzretrieval in Hypertext; Retrieval in elektronischen Enzyklopädien über Hypertext; Möglichkeiten von Hypertext für On-line-Training; Hypertext und Freitextretrieval; Möglichkeiten der automatischen Verknüpfung durch Clusteranalysen; Update von Verknüpfungen durch Häufigkeitsinformationen; integriertes Datenbankmodell für "browsing" und "query"; Retrieval unter Ungewißheitsbedingungen bei Hypertext; Hypertextschnittstellen zu relationalen Datenbanken; ...

Dabei sind vor allem in der Anfangsphase von Hypertext eher Abgrenzungsversuche unternommen worden, so wenn sich Ted Nelson gegen eine informationswissenschaftliche Vereinbarung durch die Information-Retrieval-Weit wendet (Nelson, 1973, 442), da die Konventionalität der dort zum Einsatz kommenden bibliographischen Systeme kaum dem entspräche, was einmal Vannevar Bush (1945) mit dem weitsichtigen Entwurf des Memex vorgeschwebt hatte. Wir wollen diese Abgrenzungsstrategie die *exklusive Differenzthese* nennen. Hierbei geht man davon aus, daß Hypertext und Information Retrieval zwei grundsätzlich verschiedene Paradigmen zugrundeliegen: zum einen das sogenannte Browsing- oder auch Navigationsparadigma, zum anderen das Matching- oder Such-Paradigma.

## 2. Browsing- und Matching-Paradigma

Wenn wir im folgenden beide "Paradigmen" kurz skizzieren, dann wird sich zeigen, daß die exklusive Differenzthese kaum haltbar ist. Entsprechend geht heute die Mehrzahl der Forscher von einer *komplementären Differenzthese* aus, nach der nach wie vor Unterschiede im Hypertextnavigieren und im Retrieval gesehen, aber beide Techniken bzw. das Zusammengehen beider Techniken mit Blick auf die Lösung von Informationsproblemen für unverzichtbar angesehen werden (vgl. Coombs 1989; Dunlop 1990; Ellis 1991; Knuth/Brush 1990; Rabitti/Savino; Ray/Driscoll 1990).

Peter Brown hat schon in einer Arbeit aus dem Jahr 1988 den Unterschied zwischen Browsing und Suche deutlich zu machen versucht, den Dario Lucarella 1990 aufgreift: "browsing can be characterized as going from *where to what* (presumably you know where you are in the database and you want to know what is there) whereas searching can be characterized as going from *what to where* (presumably you know what you want and you wish to find where in the database it is)." (Lucarella 1990, 83). Das eine, *what to where*, kann auch explizites Retrieval, entsprechend dem Matching- oder dem Inferenz-Paradigma, das andere, *where to what*, kann

1. Z.B. Agosti (1988); Agosti et al. (1989); Brown (1988); Campagnoni/Ehrlich (1988/1989); Coombs (1989); Cove/Walsh (1988); Croft/Turtile (1989); Crouch/Crouch/Andreas (1989); Dunlop (1990); Egan et al. (1989a, b); Ellis (1991); Frisze (1988); Frisze/Cousins (1989); Fuhr (1990); Hamnwohner (1990); Knuth/Brush (1990); Lucarella (1990); Marchetti/Belkin (1991); Marchionini/Sheideman (1988); McClelland (1989a, b); McKnight/Dilow/Richardson (1989/90); Omorato (1990); Percival/MacMorrow (1989); Pollard (1990); Smith, K.E. (1988).

assoziatives Retrieval, entsprechend dem Browsing- oder Navigations-Paradigma genannt werden.

**Browsing, Navigation.** In der Regel wird in der Informationswissenschaft der pragmatische Aspekt, also die Handlungsrelevanz von Information, herausgestellt. Information, so könnte unterstellt werden, wäre danach immer schon zielgerichtet bzw. abhängig von einem konkreten Ziel. Die Umgangssprache gibt aber ein schönes Beispiel dafür, wie "Information" auch anders gebraucht werden kann. Ein Satz wie der folgende könnte benutzt werden, wenn man die freundliche Unterstützung eines Bibliothekars, der sieht, wie man unsicher durch die Bestände der Bibliothek irt, nicht annehmen will, sei es, weil man nicht zugeben mag, daß man seinen Suchwunsch noch nicht präzisieren kann, oder sei es, weil man die Situation des Erkundens alleine genießen will:

Ich suche nichts Bestimmtes, ich informiere mich bloß.

Die Suche nach Information kann also durchaus nur sehr vage zielgerichtet sein. Entsprechend kann "sich informieren" auch "schmökern" oder "sich unverbindlich umsehen" bedeuten, allerdings schon im Kontext eines noch vagen Informationsbedürfnisses. Damit ist "Browsing" angesprochen. Wir wollen dabei die folgenden Formen von Browsing unterscheiden<sup>2</sup>:

a) **Gerichtetes "Browsing" mit Mitnahmeeffekt.** In benutzerfreundlichen, systematisch strukturierten und frei zugänglichen Bibliotheken wird diese Form von Browsing dadurch unterstützt, daß sachlich verwandte Publikationen auch räumlich benachbart in Gruppen zusammenstehen. Auch wenn die einzelnen Einheiten innerhalb dieser Gruppen meistens nach linearen (numerischen oder alpha-numerischen) Prinzipien geordnet sind, also relativ leicht individuell gefunden werden können, so stellt sich für den Endnutzer die systematische Gruppe zunächst als thematische Einheit dar, bei der jedes einzelne Element potentiell relevant ist. In diesem Sinne bedeutet "Browsing" also nicht das beliebige Durchwandern einer Bibliothek, sondern das Stöbern in einer schon vor-selektierten Menge. Die systematische Gruppierung ermöglicht "Browsing" als Mitnahme-Effekt. Er tritt dann ein, wenn man bei der Suche nach bestimmten Büchern auf thematisch verwandte Bücher trifft, nach denen man nicht gezielt gesucht hat. z.B. weil man sie nicht kannte, die aber dennoch interessant und einschlägig sind. In der Regel verfolgt man seinen ursprünglichen Plan, zu einem bestimmten Buch zu gelangen, weiter, nimmt aber zusätzliche Information mit. Von gerichtetem "Browsing" spricht man in diesem Fall, weil die ursprüngliche Richtung, die Ausrichtung auf ein Ziel, gegeben war.

b) **Gerichtetes Browsing mit "Serendipity"-Effekt.** Zuweilen geschieht es, daß man auf der Suche nach einer bestimmten Information von einer anderen Information so "beschlagnahmt" wird, daß über deren aktueller Dominanz das ursprüngliche Ziel irrelevant oder vergessen wird. Dies bezeichnet man als "*Serendipity*-Effekt". Hierbei wird die Grenze zwischen chaotischem Informationsverhalten und kreativer Aufnahme von Information fließend. Auch dieses "Browsing" ist zumindest zu Beginn gerichtet, also nicht vollkommen frei initiiert, dann aber eher frei assoziierend.

<sup>2</sup> Ausführlicher in Kuhlén (1991a) 128ff

c) **Ungerichtetes "Browsing".** Von "Browsing"-Effekten kann man aber auch sprechen, wenn noch gar kein konkreter Plan, ein bestimmte Objekt zu finden, besteht. Es ist also keine Richtung vorgegeben. Dies ist typisch für eine Situation, die in der Informationswissenschaft als "anomalous state of knowledge"<sup>3</sup> bezeichnet wird, also ein Zustand, bei der einer Person zwar bewußt ist, daß sie zur Lösung eines Problems Information benötigt, aber noch nicht genau, welche. Das "Browsing" ist erfolgreich, wenn, zumindest nach der subjektiven Einschätzung, derart befriedigende Information gefunden wird, daß das den "anomalous state of knowledge" verursachende Problem gelöst werden kann, bzw. wenn man den Eindruck hat, daß die gefundene Information in die richtige Richtung weist.

d) **Assoziatives "Browsing".** Nicht mehr initiiert von einem deutlichen oder vagen Informationsproblem ist diese Form des Browsing. Vielmehr werden in einer Situation, z.B. beim Lesen eines Textes, bei der Ansicht einer Graphik oder bewegter Bilder, beim Anhören von Tonmaterial, bestimmte Aspekte - Wörter, Details einer Graphik oder einer Szene - aus welchen Gründen auch immer so attraktiv, daß man zu ihnen mehr wissen will<sup>4</sup>. Überläßt man sich erst einmal dieser freien Assoziation, dann können durchaus längere Assoziationsketten auftreten. Es tritt dann der typische Effekt auf, daß nicht mehr gezielt nach Information gesucht wird, sondern man sich vom Reizangebot treiben läßt, bis kein starker Anreiz mehr vorhanden ist, weiteren Angeboten zu folgen<sup>5</sup>.

**Matching, Inferenz.** Im klassischen Boole'schen Retrieval dominiert das sogenannte Matching-Paradigma. Es geht letztlich von der Annahme aus, daß Benutzer in der Lage sind, ihr Suchproblem in konzeptorientierten Frageformulierungen auszudrücken, die dann mit ebenfalls konzeptorientierten Suchdateien/Indices, die als Ergebnis gleichermaßen konzeptorientierter Inhaltserschließung der Ausgangsdokumente entstanden sind, verglichen und, im positiven Fall, "gematcht" werden. Wir haben oben mit der These des "anomalous state of knowledge" angedeutet, daß die Grundannahmen einer "rationalen" Matching-Suche nicht unbedingt mit den psychischen Ungewissheitsfaktoren korrespondieren. Entsprechend wird auch im Kontext des Information Retrieval freieren Formen der Suche das Wort geredet, z.B. durch das von Bates (1986) auf exemplarische Weise formulierte explorative Paradigma oder durch Vorschläge zum "Passagen"- bzw. Absatz-Retrieval (Al-Hawamdeh/Smith/Willet 1991). Die in der Informationspraxis, d.h. beim Online-Retrieval des Informationsmarktes, nach wie vor vorherrschende Dominanz des Boole'schen Retrieval ist, entsprechend den Ergebnissen der Forschung, keineswegs dem Information Retrieval inhärent. Entsprechend

<sup>3</sup> Vgl. Brooks et al. 1986, 39: "Users may be able to recognize that their own knowledge with respect to some problem is anomalous, and will be able to describe the problem itself, but are usually unable to say precisely what is required to resolve the anomaly".

<sup>4</sup> Cove/Walsh (1988) nennen dies "serendipity browsing", sehen den "Serendipity"-Effekt als allein durch Assoziationen verursacht, die bei der Rezeption, z.B. von On-line-Materialien, entstehen.

<sup>5</sup> Am Ende einer Assoziationskette läßt die Aufmerksamkeitsspanne nach, es tritt leicht ein fatales Verlorenheitsgefühl auf, in der Hypertext-Metaphorik häufig als "lost in hyperspace" bezeichnet. Assoziationsketten haben offenbar die Eigenschaft, daß man sich ihnen so lange anvertraut, wie man sie in einem konzeptuellen (vielleicht emotionalen) Kontext stellen kann. In der Regel wird man sich nicht abbrechen, solange die weiteren Assoziationen Sinn machen, mit dem Risiko, daß bei längeren Ketten der Ausgangspunkt kaum noch rekonstruierbar ist.

<sup>6</sup> Beim Passagen-Retrieval werden nicht Texte, sondern verwandte Abschnitte verknüpft. Dies ist auch die Grundidee des "dynamic book" (Weyer 1982), eines der Vorläufer heutiger Hypertextsysteme. Im "dynamic book" sind ebenfalls alle Teile durch logische Suchstrukturen, die über leistungsstarke Register operieren (häufig umfangreicher als die Texte selber), direkt ansprechbar. Darüberhinaus können die Benutzer selber weitere Inhaltsbeschreibungen (potentielle Suchterme) und Verknüpfungen anbringen. Auch experimentelle Systeme wie "Eyebrows" (Cove/Walsh 1988) oder IOTA (DeFude/Chiamarella 1987) erlauben "Browsing" bzw. Exploration in Volltexten, z.B. unter Ausnutzung von Wort-Ko-Occurrenzen über Kontextoperatoren.

spricht man bei den vektor-orientierten (Prototyp: SMART), probabilistischen oder clusterbasierten Retrievalmodellen auch eher von "partial match retrieval".

In neueren Retrieval-Ansätzen (van Rijsbergen 1986, Croft/Lucia/Cohen 1988, Fuhr - in diesem Band) wird entsprechend der Auflösung des Matching-Ansatzes das Retrieval redefiniert als Prozeß plausibler Inferenzen, sei es auf der Basis einer Nicht-Standard-Logik oder eines Bayes'schen Inferenznetzes: "In this approach, the retrieval process is regarded as the problem of determining an implication relationship between a document and a query and assessing the plausibility of that implication. Thus retrieval is viewed as a process of inference and each selection of a collection item in response to a query can be treated as a theorem to be proven or disproven on the basis of the interpretation of the assertions for the domain of interest" (Lucarella 1990, 83f).

### 3. Informationelle Mehrwerte

Weitergehend als die komplementäre Differenzthese ist die hier vertretene Mehrwertthese. In einem Beitrag für die ISI-Konferenz 1991 haben ich einen ersten Ansatz zur Theorie informationeller Mehrwerte unternommen und habe dabei systematisch die folgenden vier Typen informationeller Mehrwertbildung unterschieden (Kuhlen 1991b):

1. Größerer Informationswert der elektronischen Version gegenüber der korrespondierenden konventionellen
2. Größerer Informationswert durch Verbesserung einzelner Komponenten eines elektronischen Produkts (oder einer Dienstleistung) oder durch Verbesserung der Gesamtleistung (auch inhärenter Mehrwert genannt)
3. Größerer Informationswert durch Agglomeration ursprünglich isolierter Leistungen oder Produkte (auch Mehrwert durch Integration genannt)
4. Kombination verschiedener Typen von Informationsprodukten oder -dienstleistungen

Ich möchte im folgenden kurz auf jeweils ein Beispiel aus dem zweiten und dritten Mehrwerttyp eingehen. Beim ersten wird darauf hingewiesen, wie eine traditionelle Technik aus dem weiteren Information Retrieval, das Abstracting, für Hypertext mehrwertbildend eingesetzt werden kann. Beim zweiten wird auf den Mehrwerteffekt durch Kombination von Hypertext und Online-Retrieval verwiesen.

#### 3.1 Mehrwerteffekte durch die Verwendung von Zusammenfassungen bei Hypertext

Zusammenfassungen fügen Texten einen *informationellen Mehrwert* zu, der nach vorherrschender Meinung der Informationswissenschaftler (vgl. Kuhlen 1989) im wesentlichen in der *Referenzfunktion* gesehen wird. Zusammenfassungen sollen demnach ausreichende Informa-

tionen darüber liefern, ob es sinnvoll ist, sich den Volltext anzuschauen (oder auch nicht). Zuweilen - im Falle der informativen Referate - können diese die Lektüre der Volltexte sogar überflüssig machen. Sie haben dann nicht mehr bloß indikative, referenzierende, sondern durchaus auch substituierende Funktion. Zusammenfassungen können als lineare Texte, aber auch in strukturierter Form (als Tabellen, Datensätze oder Framemuster) dargestellt werden. Zusammenfassungen können über die Referenz- und Informationsfunktion hinaus durchaus auch als *kognitiv relevant* angesehen werden, und zwar allgemein zur Bildung von Vorurteilen und dann spezieller, mit Blick auf deren Einsatz bei Hypertext, als Mittel zur Orientierung in komplexen oder auch nur großen Hypertextbasen.

Hypertext, zunächst wegen des durchgängig verwirklichten Prinzip der direkten Manipulation ein scheinbar leichtes, d.h. benutzerfreundliches Medium, erweist sich angesichts größer werdender Hypertextbasen als durchaus kompliziertes Medium. Dies gilt allemale für den Aufbau von Hypertextbasen (angefangen von der Modellierung, der Konversion - bei der Übernahme bestehender Objekte, wie Texte, in die Hypertextbasis -, der inhaltlichen Beschreibung der informationellen Einheiten, der Realisierung und Kontrolle der Verknüpfungen und der Bereitstellung von Metainformationen, wie Register, Glossare), aber auch für die Benutzung fertiger Hypertextbasen selber. Rein assoziative Hypertexte - darunter verstehen wir *Hypertextsysteme der ersten Generation*, die die Verknüpfungen zwischen Hypertexteinheiten oder Teile von ihnen auf rein assoziative Art realisieren, also die Verknüpfungen nicht weiter semantisch spezifizieren - führen den Benutzer leicht zu einer Desorientierung. Der Gewinn an informationellen Mehrwert, der bei Hypertext häufig darin gesehen wird, daß man nicht eingefahrenen Gleisen folgt, also nicht feste und gesicherte Antworten auf Fragen sucht, die man genau formulieren kann, sondern durch das Assoziieren immer neue Anregungen bekommt und das Ausgangsproblem auf kreative Weise in immer neue Kontexte stellen kann, droht dann verlorenzugehen.

Orientierungsprobleme werden - analog den bekannten Nivergel'schen Kontrollfragen zur Steuerung des Design von Mensch-Maschine-Schnittstellen allgemein - häufig über Fragen, wie den folgenden angesprochen:

Wo bin ich?  
Wie kam ich hierher?  
Warum kam ich hierher?  
Habe ich alles Relevante schon gesehen?  
Wohin kann ich, sollte ich gehen?  
Was erwartet mich; ist es wert, daß ich dem Angebot nachgehe?

Wir wollen uns kurz mit dem Komplex beschäftigen, der mit der letzten Frage angesprochen ist. Es soll dargestellt werden, welchen Beitrag Techniken des Abstracting, oder neutraler ausgedrückt: der orientierenden Überblicksinformation in diesem für die Akzeptanz von Hypertext zentralen Teilbereich der Orientierungsproblematik leisten können.

Hypertextbasen können im wesentlichen über vier Ansätze übersichtlicher gemacht werden (vgl. Kuhlen 1991a, 124):

- a) durch Strukturierung der gesamten Hypertextbasis, z.B. durch durchgängige Hierarchisierung oder durch die Verwendung anderer Modellierungstechniken, z.B. der Objekttypenmethode;
- b) durch Orientierungsformen wie Register, Inhaltsverzeichnisse die allgemein unter dem Begriff der Metainformation zusammengefaßt werden; dazu gehören auch die Zusammenfassungen;
- c) durch Navigationsangebote wie autoredefinierte Pfade, durch die Teilmengen von Hypertexteinheiten unter thematischen Gesichtspunkten zusammengefaßt werden sind, denen die Benutzer nur einfach folgen müssen (wohl mit der Freiheit, diese Pfade jederzeit bei attraktiven Angeboten verlassen zu können);
- d) durch pragmatisches Design von Hypertextbasen, indem das System von sich aus, z.B. durch Schließen aus Information über den Benutzer, seine Situation oder sein bisheriges Dialogverhalten, weitere Nutzungsangebote macht.

Gehören die ersten drei Techniken zur gängigen Ausstattung von Hypertextsystemen der zweiten Generation, so sind pragmatischen Komponenten erst unter Verwendung wissenschaftlicher Verfahren, wie Benutzermodellierung oder Steuerung durch Dialoghistorien, in Systemen der dritten Generation zu verwirklichen.

Zusammenfassungen gehören zweifellos zu den Metainformationsmöglichkeiten. Sie gehören als zunächst zu der Ausstattung der Zweite-Generation-Systeme. Wie wir am Ende des Beitrags jedoch andeuten wollen, müssen zur Erzeugung flexibler benutzerangepaßter Zusammenfassungen wissenschaftliche Verfahren eingesetzt werden, so daß sie dann auch zu den Verfahren unter (d) zu rechnen sind.

Zusammenfassungen können, wie angedeutet, für einzelne Einheiten, für größere Einheiten wie Pfade oder auch für ganze Hypertextbasen eingesetzt werden. Sinn macht die Verwendung von Zusammenfassungen für einzelne Hypertexteinheiten nur, wenn anders als bei kartorientierten Systemen (bei denen also die vollständige, auf eine Kartengröße reduzierte Information auf dem Bildschirm schon sichtbar ist, also nicht mehr gesondert angekündigt werden muß) die Hypertexteinheiten als rollende Felder organisiert sind, also größere Texte aufnehmen können. Dies hängt in erster Linie von der Textsorte bzw. dem Domänenbereich ab. Zusammenfassungen können auch zur Übersicht über vordefinierte Pfade (Zellweger 1989) oder über komplexe Hypertexteinheiten (*composite nodes* im Sinne Hatalasz 1987/88) verwendet werden.

Zusammenfassungen in Hypertext können als selbständige Objekte den informativen Teilen (Hypertexteinheiten, Pfaden etc.) vorgeordnet werden, so daß erst eine Bestätigung zum Öffnen des (Voll-)Textfensters führt, so bei HyperTies (vgl. Shneiderman/Kearsley 1989), oder aber sie werden parallel, aber in getrennten Fenster angezeigt.

Zusammenfassungen haben die wichtige Funktion, durch Vorabinformation eine *Erwartungshaltung* (ein vorläufiges kognitives Netz) aufzubauen, das die Rezeption der möglicherweise folgenden konkreteren Information wesentlich erleichtert. Sie können damit im Sinne der

Hermeneutik (Gadamer) als kognitiv nützliche *Vorurteile* angesehen werden oder auch - im Hypertextkontext - als wichtiges Element einer "rhetoric of arrival" (Landow 1987).

Textuelle oder strukturierte Zusammenfassungen unterscheiden sich prinzipiell nicht in den Zwecken der Relevanzentscheidung oder der kognitiven Unterstützung. Signifikante Ausdrücke in Zusammenfassungen sind gut als Ausgangspunkte (Anker) für Verknüpfungen zu anderen Einheiten geeignet. Diese ist bei textuellen Zusammenfassungen möglich, aber vor allem bei *strukturierten Zusammenfassungen* können die Einträge können dazu verwendet werden, einen selektiven ("zoomenden") Zugriff auf Details oder Aspekte der informativen Teile von Einheiten zu ermöglichen. Strukturierte Referenzteile sind unter dem Einfluß von Wissensrepräsentationstechniken häufig in Form von "Frames" oder Skripts organisiert. Dies geschieht unter der Annahme, daß es möglich ist, Sachverhalte weitgehend vorstrukturiert, entsprechend ihren Eigenschaften, zu beschreiben. Dies geschieht z.B. bei "Frames" durch die Festlegung von Eigenschaften von Konzepten in "Slots". Die jeweils konkrete Information kann dann über die Aktivierung der realen "Slot"-Einträge eingesehen werden.

Strukturierte Zusammenfassungen sind vor allem dann attraktiv, wenn sie durchgängig in der Hypertextbasis nach *gleichen Aufbauprinzipien* organisiert sind, also z.B. in ähnlicher Funktion gleiche "Frame"- oder Skript-Strukturen aufweisen. Boyle/Snell (1989/90) stellen am Objektbereich der Tiermedizin ein wissensbasiertes, unter Anwendung von HyperCard entwickeltes Hypertextsystem vor, das - in Weiterführung der von Malone (1987) vorgeschlagenen Idee der Verwendung halbstrukturierte Muster ("Templates") oder "Frames" zur Erschließung von Texten in elektronischen Postsystemen - die Hypertexteinheiten nach vor-spezifizierten (insgesamt sieben) Feldern und zusätzlich unstrukturiertem Text gliedert. Dies ist nützlich als Merkposten beim Aufbau von Hypertextbasen, vor allem aber erleichtert die Vorstrukturierung die in Hypertextbasen häufig gefährdete Orientierung, da vergleichbare Information an der gleichen Stelle zu finden ist. Auch Neuwirth/Kaufer (1989) realisieren Zusammenfassungen als Instantiierungen eines strukturell vorgegebenen Schemas, das Automatisch als Anregung zum Aufbau von Zusammenfassungen dient. Das vorgegebene Schema besteht aus einem Hauptpfad ("main path"), dessen einzelne Einheiten etikettiert sind, z.B. "see the issue", "choose solution" oder "define problem". In Ergänzung zu dem Hauptpfad gibt es Pfade, die darüber informieren, welche Positionen der Autor opponiert ("faulty paths") und schließlich führen Pfade wieder zum Pfad zurück ("return paths"). Bei realen Instantiierungen werden diese formalen Vorgaben mit Informationen aus den vollständigen Einheiten zum Zwecke eines Überblicks und zum Erkennen der argumentativen Verknüpfungen zwischen einzelnen Pfadelementen gefüllt.

Zusammenfassungen können auch auf *graphischer Basis* realisiert werden. Aus der Informationswissenschaft gibt es dazu ein früher Beispiel mit den *document maps* von Doyle (1961), mit denen das Wissen ganzer Bibliotheken in konzeptuellen Netzwerken und Dokumenten(land)karten zugriffsfähig gemacht werden sollte. Graphisches Abstracting wurde auch in den siebziger Jahren an der Ohio State University durchgeführt. Dazu hat Strong (1974) einen Vorschlag vorgelegt, bei dem auf der Grundlage der Fillmore'schen Kasusgrammatik, angereichert durch semantische Merkmale, die zentralen Konzepte von Texten in Textwissensgraphen vernetzt werden. Als Stilmittel werden dabei vor allem unterschiedlich graphisch gestaltete Kanten eingesetzt. Als graphische Zusammenfassungen sind auch die im TOPIC-

Projekt vorgeschlagenen Textgraphen anzusehen (Hahn/Reimer 1985, 1988; Hahn et al. 1990). Als hypertextgemäße Form von graphischen Zusammenfassungen wurde anfänglich globale graphische Übersichten angesehen (vgl. Conklin 1987). Nach den kritischen Untersuchungen von Bernstein (1988) und Utting/Yankelovich (1989) dürfte jedoch gesichert sein, daß als graphische Übersichten bzw. Zusammenfassungen allenfalls *local browser* sinnvoll sind. Auch graphische "Browser" erfüllen die kognitive Funktion der Orientierung durch übersichtliche Vorabinformation.

Besonders attraktiv sind Zusammenfassungen dann, wenn sie *flexibel* an unterschiedliche Benutzerbedürfnisse angepaßt werden können. Dies ist angesichts des hohen intellektuellen Aufwandes bei konventionellen Systemen kaum möglich, sondern erfordert Techniken des automatischen Abstracting, und zwar auf wissensbasierter Grundlage. Dies wurde im System TWRM-TOPOGRAPHIC realisiert (Sonnenberger 1988; Kuhlén et al. 1989; Kuhlén 1991a), wobei in *question time* durch den Vergleich einer Benutzeranfrage mit den im System vorhandenen Textwissenstrukturen Zusammenfassungen unterschiedlicher Struktur generiert werden konnten. Zusammenfassungen sind dort eine von verschiedenen *Kaskadierungsstufen*, auf denen sich Benutzer im Hypertextstil die jeweiligen Textinformationen erarbeiten können.

### 3.2 Verbindung von Hypertext zu On-line-Informationsbanken

Durch den Einsatz von Frageverknüpfungen ("query links") (vgl. Smith, K.E. 1988; Percival/MacMorrow 1989; McClelland 1989a,b) entwickelten sich Hypertextsysteme zu offenen Kommunikationssystemen, die externes Wissen aufnehmen können (vgl. Kuhlén 1991a). Diese Komponente scheint uns für Hypertext sehr wichtig zu sein, denn die Navigation in schon vorliegenden Materialien, wie Handbücher, ist zwar reizvoll genug - eine ganz besondere und wichtige Herausforderung ist aber die laufende Einbettung neuer Materialien in bestehende.

Nach unserer Einschätzung sollte bei dem Angebot von Frageverknüpfungen zwischen unerfahrenen und geübten Nutzern unterschieden werden:

a) Im ersten Fall könnte es sich um Systemangebote handeln, die beim Aufbau einer Hypertextbasis von Systementwicklern an solchen Stellen vorab eingerichtet werden, von denen sie annehmen, daß in kurzen Zeitabständen neue Literatur produziert wird. Frageverknüpfungen sind dann in der Leistung sogenannten SDI-Diensten vergleichbar, bei denen ein automatisiertes Informationssystem seinen Benutzern in festen Zeitabständen auf der Basis eingespeicherter Suchprofile die neueste Literatur nachweist. Bezüglich Hypertext bedeutet dies, daß der Benutzer beim Navigieren durch die Hypertextbasis an besonders gekennzeichneten Stellen, z.B. über spezielle, graphisch gekennzeichnete Frageanzeiger, darauf hingewiesen wird, daß er zu dem an der Stelle angesprochenen Thema neuere Literatur einholen kann. Voraussetzung ist natürlich, daß das Hypertextsystem über entsprechende Telekommunikationsmöglichkeiten verfügt. Der Benutzer selber muß darüber im Prinzip nichts wissen, auch nichts über Techniken der Datenbankrecherche (vgl. Kesselman/Trapasso

1988). Als Ergebnis der Aktivierung der Frageverknüpfung wird dem Benutzer der aktuellen Stand der über die Verknüpfung verbundenen On-line-Informationsbanken angezeigt.

b) Für Benutzer mit Erfahrung beim Umgang mit externen Datenbanken wäre eine Version erwünscht, mit der der Benutzer im Ausgang von einer nach weiterer Information verlangenden Passage in der Hypertextbasis von sich aus, auf der Basis seines Wissens, eine On-line-Recherche durchführen kann. Systemhilfen, wie Auflistung potentiell einschlägiger Datenbanken, Hinweise auf Kosten, Unterstützung der Frageformulierung in terminologischer und syntaktischer Hinsicht, wie sie zum Teil heute schon in kommerziellen "Gateway"-Systemen realisiert sind<sup>7</sup>, sind dabei erwünscht. Bei erfolgreicher Recherche kann der Benutzer an der entsprechenden Stelle in seiner Hypertextbasis eine Frageanzeige vermerken, die er - entsprechend (a) - zu einem späteren Zeitpunkt ohne weiteren Aufwand aktivieren kann, um sich den neuesten Stand in den jeweiligen Datenbanken anzeigen zu lassen.

In beiden Fällen sollte das Hypertextsystem Möglichkeiten des Einspeicherns relevanter Dokumente bzw. -passagen aus externen Beständen vorsehen ("down-loading") und weiterhin Möglichkeiten, diese neuen Texte bzw. fragmentierten Einheiten in das existierende Verknüpfungsgeflecht der bisherigen Hypertextbasis einzubinden.

### 3.3 Systemdesign von WITH

Diese Lösung dieser Aufgabe - automatische Segmentierung neuer, über On-line-Recherchen eingespeicherter Dokumente in hypertextgerechte, d.h. kohäsiiv geschlossene informationelle Einheiten und ihre automatische Verknüpfung - ist Forschungsziel des in der Konstanzer Informationswissenschaft in Entwicklung befindlichen Systems WITH (wissensbasiertes, textorientiertes Hypertextsystem). In den folgenden beiden Abbildungen wird zunächst das Systemdesign von WITH als *offenes Hypertextsystem* gezeigt, dessen Hypertextbasis sich kontinuierlich aus heterogenen externen Texten aufbaut und danach die Systemumgebung von WITH, die über verschiedene Parsing-Ebenen operiert. Die Analyse der Volltexte beginnt mit der formalen Segmentierung durch einen Text-Parser, der SGML-Textauszeichnungen ausnutzt. Die so zunächst noch in formale Einheiten zerlegten Texte werden danach (in den ersten Phasen des Projektes durch intellektuelle Analyse) semantisch auf eine angereicherte Frame-Sprache abgebildet und entsprechend semantisch stimmig in Hypertexteinheiten eingeteilt. Dazu können die auf der ersten Stufe nur formal bestimmten Einheiten nach semanti-

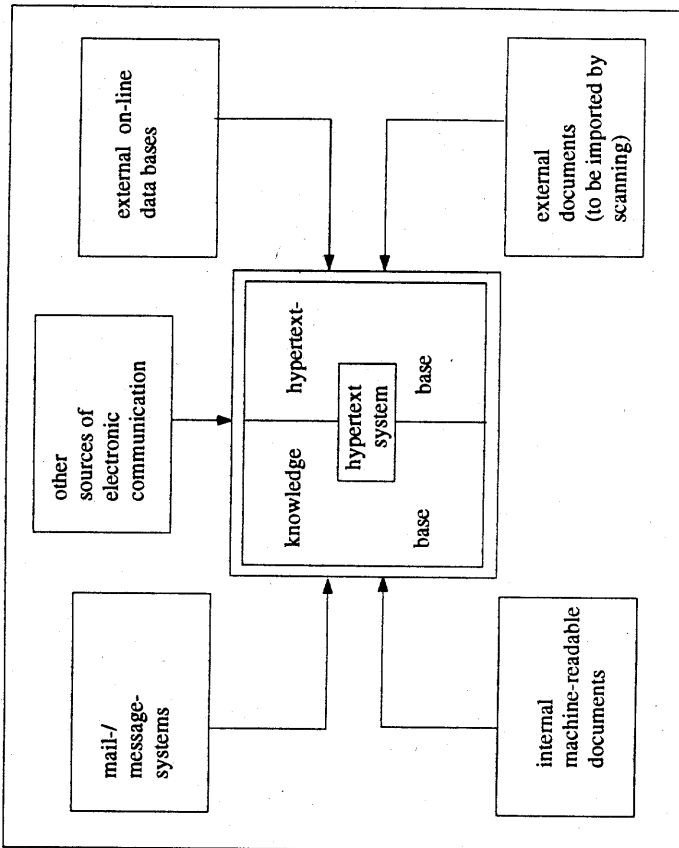
<sup>7</sup> Mit "Gateways" bezeichnet man im Kontext des Information Retrieval bzw. des Informationsmarktes allgemein Systeme, die Dienstleistungen zur Verbesserung des Zugriffs auf On-line-Informationsbanken gestatten. "Gateways" machen für Benutzer auch die Nutzung vielfältiger Datenbanken in unterschiedlichen Systemen möglich, ohne daß der Benutzer individuelle Nutzungserträge abschließen muß. "Gateways" sind deutliche Hinweise dafür, daß sich der On-line-Informationsmarkt auch für unerfahrene, nicht-professionelle Nutzer öffnet - eine Tendenz, die mit der Entwicklung von Hypertext zusammenhängt.

Eine kommerzielle Anwendung der Verknüpfungsidee von Hypertext im On-line-Retrieval wird von Nesbit (1990) beschrieben. Seit Mai 1989 bietet "BRS Information Technologies" einen sogenannten BRS/Link an, mit dem ein Eintrag aus einer bibliographischen Datenbank (z.B. MEDLINE) mit der fachlich entsprechenden Volltextdatenbank (bislang Comprehensive Core Medical Library - CCML) verbunden werden kann, ohne daß eine neue Recherche aktiviert werden muß. Allerdings können bislang Abbildungen und Tabellen, also graphische Informationen, nicht eingesehen werden. Dies ist bei einer Ausweitung der CD-ROM-Produkte aber bald zu erwarten. BRS experimentiert zur Zeit - nach Nesbit (1990) - mit weiteren Datenbasen, um Hypertextmerkmale, auch zur Unterstützung der Frageformulierung, einzuführen (vgl. Marchetti/Belkin 1991).

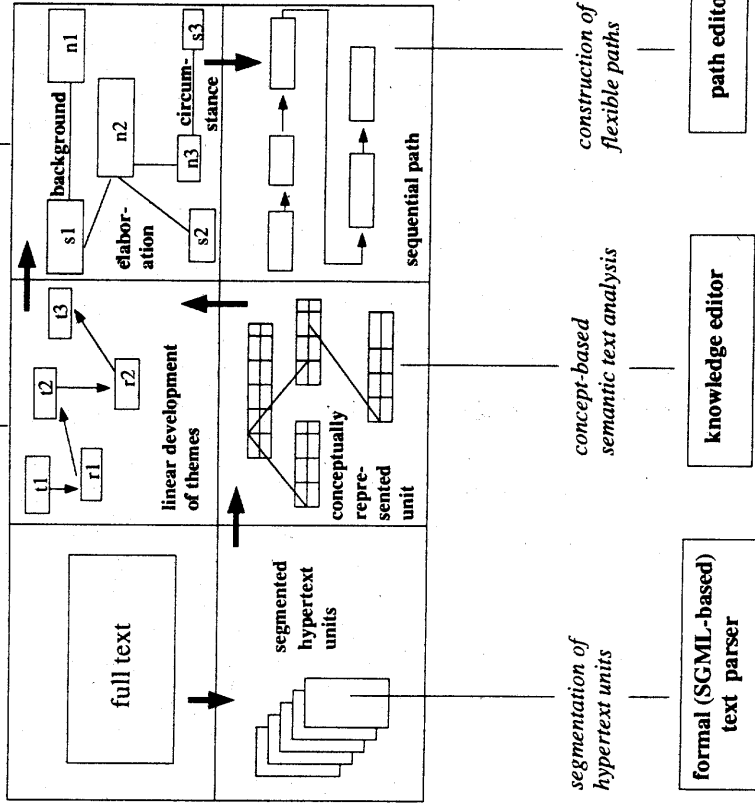
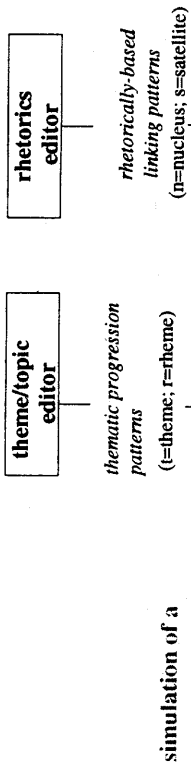
# WITH

Knowledge representation and retrieval of information in a text-based hypertext system

objective: integration of text-based knowledge units from heterogeneous sources to support information-intensive knowledge workers



System desing of WITH



System environment of WITH

