

Universität Konstanz
Fachgruppe Politik-/Verwaltungswissenschaft
Informationswissenschaft
Projekt Informationsvermittlung
Projekt TOPOGRAPHIC
Postfach 5560
7750 Konstanz 1

Projekt Informationsvermittlung
Teilprojekt TOPIC
&
Projekt TOPOGRAPHIC
Projektleiter: Prof. Dr. Rainer Kuhlen

Gefördert durch den Bundesminister für
Forschung und Technologie

Projektträger:
Gesellschaft für Information und Dokumentation (GID)
PT 200.08 & PT 10200.15

TOPIC II / TOPOGRAPHIC II

Automatische Textkondensierung und
text-orientiertes Informationsmanagement

- Projektziele -
- State-of-the-Art -

U. Hahn / R. Hammwöhner /
R. Kuhlen / U. Reimer / U. Thiel

Bericht TOPIC-12/84 & TOPOGRAPHIC-3/84

Dezember 1984

Dieser Bericht ist ein Extrakt aus der Anlage zum Projektantrag
"Automatische Textkondensierung und text-orientiertes Informa-
tionsmanagement" vom Oktober 1984.

INHALT

1. Gesamtziel des Vorhabens	1
2. Überblick über die Arbeitsziele des Vorhabens	5
3. Ausführliche Beschreibung der Arbeitsziele des Vorhabens	10
3.1 Text-Parsing	11
3.2 Textrepräsentation	14
3.3 Texttransformation	17
3.4. Text-orientiertes Informationsmanagement	18
4. State-of-the-Art	28
4.1 Übersicht über Methoden der Volltextverarbeitung in in Informationssystemen (U. Hahn, U. Thiel)	28
4.1.1 Der informatische Ansatz	28
4.1.2 Der statistische Ansatz	30
4.1.3 Der linguistische Ansatz	30
4.1.4 Der wissensbasierte Ansatz	32
4.1.5 Der kognitive Ansatz	34
4.2 Ein integrales Konzept der Textanalyse und des Text- informationsmanagement in einem wissensbasierten Volltext-Informationssystem	41
4.2.1 Textanalyse in TOPIC (U. Hahn, U. Reimer)	41
4.2.1.1 Text-Parsing	41
4.2.1.2 Textrepräsentation	47
4.2.1.3 Textkondensierung	55
4.2.2 Textinformationsmanagement in TOPOGRAPHIC (R. Hammwöhner, U. Thiel)	58
4.2.2.1 Formen graphischer Interaktion	58
4.2.2.2 Alternative Formen des Retrieval	63
5. Kurzberichte zu den Vorläuferprojekten	67
5.1 Kurzbericht zum Projekt TOPIC I	67
5.2 Kurzbericht zum Projekt TOPOGRAPHIC I	67
5.3 Veröffentlichungen aus den Projekten TOPIC und TOPOGRAPHIC	70

Vorwort

Der vorliegende Bericht ist aus Fragmenten des Projektantrags zusammengesetzt, mit dem die Arbeiten an den Projekten TOPIC und TOPOGRAPHIC am Lehrstuhl für Informationswissenschaft an der Universität Konstanz fortgesetzt werden sollen.

Mit der Herausgabe dieser Fragmente in Berichtsform werden zwei Anliegen verfolgt:

1. Die Information der Fachwelt über die Ziele der Projektfortsetzung, die mit TOPIC II / TOPOGRAPHIC II verfolgt werden (Kap. 1-3)
2. Die Veröffentlichung der umfangreichen Aufarbeitung der Literatur in diesem relativ diffusen Überschneidungsbereich unterschiedlicher Disziplinen, die wohl insbesondere für diesen Sektor der konstruktiv-experimentellen Informationswissenschaft zugleich auch einer inhaltlichen und methodischen Standortbestimmung dient (Kap. 4)

1. Gesamtziel des Vorhabens

Automatische Textkondensierung und text-orientiertes Informationsmanagement

(TOPIC II / TOPOGRAPHIC II)

Mit TOPIC II / TOPOGRAPHIC II wird das Ziel verfolgt, die aus Mitteln des BMFT geförderten Arbeiten von TOPIC I (seit 1/82) und TOPOGRAPHIC I (seit 6/83) fortzuführen und mit der Bereitstellung eines Prototyps für automatische Textkondensierung und volltextorientiertem flexiblen Retrieval vorläufig abzuschließen.

Die Arbeiten in der zweiten Projektphase verfolgen grundsätzlich die gleiche Gesamtkonzeption, gehen aber in quantitativer (z.B. Erweiterung des Wortexperten-Bestandes) und qualitativer Hinsicht (z.B. flexibler graphikgesteuerter Retrievaldialog) über TOPIC/TOPOGRAPHIC I hinaus.

Eine wesentliche Erweiterung gegenüber den Arbeiten der ersten Phase wird darin bestehen, daß die bisherige, methodisch bedingte Beschränkung der Analyse und der Verwaltung auf jeweils einen Text aufgehoben wird. Mit der dann nötig werdenden Vernetzung der Wissensstrukturen und der Zugriffsstrategien werden die ersten Schritte in Richtung des Aufbaus einer Datenbasis getan.

Zielsetzung

Informationswissenschaftliche Forschung begründet sich wesentlich durch die Entwicklung von Verfahren und Methoden, Fachinformation nachzuweisen bzw. direkt bereitzustellen (Information Retrieval).

Wurde bislang der Bedarf nach solchen Verfahren und Methoden in erster Linie mit dem Hinweis auf die steigende Wissensproduktion in allen Fachgebieten und die ebenfalls anspruchsvoller werdenden Benutzerbedürfnisse begründet, so kommt in den letzten Jahren (und verstärkt in der Zukunft) die Tatsache hinzu, daß sich Produktion und Distribution von Wissen zunehmend auf elektronische Medien abstützen. Dazu gehören beispielsweise die computergesteuerte Satzerstellung von Zeitschriften und Monographien im Verlagswesen, die sich abzeichnende flächendeckende Verwendung von Text/Edit-Systemen in allen Organisationen der primären Wissensproduktion, der Einsatz von Post-, Botschafts- und Konferenz-Systemen in der lokalen und verteilten Fachkommunikation, der Aufbau von Volltextdatenbanken/Text-Depots mit zugeordnetem "publishing-on-demand".

Die digitale Form wird in Zukunft der normale Textzustand sein. Wir konzentrieren uns im folgenden auf die textuellen Probleme der Elektronisierung von Fachinformation.

Anders als beim gedrucktem Medium hat der Mensch zu Computerspeichern keinen direkten Zugriff und hat auch kaum Möglichkeiten, sich durch "Browsing" einen Überblick zu verschaffen.

Die genuin informationswissenschaftliche Aufgabe, die Kommunikationswege zwischen Wissensproduzenten und -nutzern offenzuhalten, stellt sich angesichts der angedeuteten Veränderungen in den Textproduktions- und Distributionsformen erneut und eindringlicher. Es wird entsprechend davon ausgegangen, daß in Ergänzung zur elektronischen Volltexterstellung und -speicherung Verfahren der Volltextverarbeitung entwickelt werden müssen, mit denen zumindest erst einmal die potentiell einschlägigen Texte nachgewiesen werden können (Referenz-Funktion).

Dabei kann man davon ausgehen, daß die schon bei Textsubstraten wie Titeln oder Referaten unzulängliche Methode des Freitextretrieval (Invertierung von weitgehend unbearbeiteten Textwörtern und Bereitstellung über einfache Dateien) bei Volltexten erst recht zu unbefriedigenden Ergebnissen führen wird. Die beiden hauptsächlichlichen Fehlerquellen sind syntaktische Fehlinterpretationen (Sichtlochkarteneffekt auch bei Verwendung von Kontextoperatoren) und unzureichende semantische Kontrollierbarkeit (z.B. Synonymie- und Homonymie-Probleme).

Bei weitergehenden Ansprüchen, die über die Referenzfunktion hinausgehen (z.B. direkter Nachweis von Textfakten, Berücksichtigung von konzeptuellen Veränderungen in der Textbasis, kognitiv plausiblere Dialog- und Präsentationsverfahren, Verwendung von Retrievalheuristiken) und bei der Ausweitung der Textsorten auf Volltexte (Fachtexte aus Zeitschriften) sind aufwendigere und textspezifische Analyse- und Präsentationsverfahren unabdingbar.

Angesichts dieser Herausforderung stellt sich das zur Förderung beantragte Projekt die Aufgabe, einen Beitrag zur Entwicklung volltextorientierter Informationssysteme zu leisten. Das Projekt steht damit in der Tradition des Information Retrieval.

Entsprechend der in der Theorie des Information Retrieval begründeten (in der Praxis der Systeme leider zumeist vernachlässigten) Einheit von Inhalterschließungs- und Retrievalkomponente hat auch der hier zu entwickelnde Systemtyp zwei Hauptteile, die aufeinander bezogen sind:

- Darstellung des Inhaltes von Texten durch textspezifische Repräsentationsstrukturen und flexibles Kondensieren
- auf Texte bezogene Verwaltung (Aufbau und Pflege) von Wissen und graphisch unterstützter Zugriff auf die Textwissensbestände

Das beantragte Forschungsvorhaben beabsichtigt, die sich aus den sechziger Jahren herleitende Forschungstradition des automatischen Abstracting mit gänzlich neuen Mitteln aufzugreifen. Die Leistungsfähigkeit des zu entwickelnden Prototyps wird in dieser Phase beschränkt sein auf - in traditionell dokumentarischen Begriffen ausgedrückt - indikatives bzw. auf Strukturreferate (Faktenwissen) ausgerichtetes Abstracting von Volltexten und entsprechendem interaktiv-graphischen Retrieval.

Der Unterschied zum traditionellen Abstracting wird darin bestehen, daß nicht versucht wird, mit der menschlichen Leistung beim Referieren zu konkurrieren, die im wesentlichen darin besteht, aus einem 10-Seiten-Text einen sinnhaften 10-Zeilen-Text zu machen, sondern daß angestrebt wird, die durch die maschinelle Verarbeitung möglich werdende Flexibilität auszunutzen. Hierfür ist das Konzept des kaskadierten Kondensierens von Texten, also des stufenweisen Präzisierens von Textinformation, entwickelt worden. Wir gehen davon aus, daß dem Benutzer der gestaffelte Zugriff zur Information, von einer sehr hohen konzeptuellen Ebene bis zur Detailinformation in den Texten selbst, möglich gemacht werden sollte. Durch entsprechende Dialogsteuerungsmechanismen kann der Grad der Spezifizierung angeboten und gewählt werden. In einer Endausbaustufe müßte das System auf einer Volltextdatenbasis aufsetzen, da über die durch die Analyseverfahren erarbeiteten Informationen (Wissensnetze) hinaus der Zugriff auf die originale Textinformation in jeder Spezifität erhalten bleiben sollte.

Um die angestrebte Leistung des zu entwickelnden Systemtyps zu erreichen, sind die folgenden Systemteile zu erarbeiten, die in Kapitel 2 und 3 näher ausgeführt werden:

1. Textanalyse (identisch mit TOPIC)
 - 1.1 Text-Parsing: Entwicklung volltextspezifischer Analyseverfahren, hier auf der Basis eines Wortexpertenansatzes
 - 1.2 Wissensrepräsentation von Texten, hier auf der Basis eines erweiterten Frame-Modells
 - 1.3 Texttransformation/-kondensation, hier auf der Basis von wissensgesteuerten Clustering-Verfahren und textlinguistischen Methoden
2. Text-orientiertes Informationsmanagement (identisch mit TOPOGRAPHIC)
 - 2.1 graphisch-interaktive Komponente zur Navigation in Wissensnetzen
 - 2.2 Entwicklung einer graphischen Retrievalsprache
 - 2.3 Anpassung der Dialogführung an die Retrievalsituation
 - 2.4 Faktenretrieval durch kaskadierten Zugriff auf Textwissen in abnehmenden Aggregierungsgraden

Angesichts des Standes der Informationswissenschaft in der Bundesrepublik kann das Projekt an keine vorhandene Forschungstradition anknüpfen. Jedoch ist die offenbar typische informationswissenschaftliche Forschungssituation gegeben, daß unter Vorgabe einer informationspraktischen Fragestellung (Volltextkondensierung mit entsprechender Retrievalkomponente) die methodischen Ansätze anderer Disziplinen aufgegriffen werden können, die funktional erfolgsversprechend erscheinen. Dazu gehören etwa: das Frame-Konzept und das Wortexpertenmodell aus der Künstlichen Intelligenz, Wissen um Kohärenz- und Kohäsionsstrukturen aus der Textlinguistik, Prinzipien der interaktiven Graphik, der Datenverwaltung und des Prototyping aus der Informatik, Prinzipien der kognitiven Ergonomie und der Benutzermodellierung aus der Kognitiven Psychologie. Und nicht zuletzt werden durch den Einsatz der MODULA-Maschine LILITH die gegenwärtigen Möglichkeiten der Terminal-/Arbeitsplatzrechner-Technologie (Windowing, mouse-Eingabe, bit-map-display, Zooming) ausgenutzt. Auch ist die gesamte Hardware-Konfiguration (Realisierung der TOPIC-Komponenten auf einem zentralen Time-Sharing-Mikrorechner und Kopplung der in TOPOGRAPHIC eingesetzten LILITH) wohl exemplarisch für die künftige Realisierung von Informationssystemen in der Verbindung von zentraler, für die Verarbeitung und Speicherung großer Datenmengen und Programmen geeigneter Rechenkapazität und lokaler, arbeitsplatzbezogener DV-Intelligenz.

In den Vorlaufprojekten TOPIC I und TOPOGRAPHIC I ist ein gewisses Ausmaß des erforderlichen Methodenwissens erarbeitet worden. Angesichts dieser Vorarbeiten wird in der beantragten Phase die Realisierung eines Prototyps angestrebt. TOPIC/TOPOGRAPHIC ist damit ein auf Anwendungsperspektiven ausgerichtetes Forschungsprojekt.

Ogleich sowohl die experimentelle Bewertung der Textkondensate als auch die Bewertung der Effekte eines interaktiv-graphischen Text-Retrieval-Konzepts Desiderate informationswissenschaftlicher Forschung sind, können sie im projektierten Zeit- und Personalrahmen nicht realisiert werden. Die erstellten Systemteile können in ein solches Vorhaben eingebracht

werden, bei dem informationelle Problemstellungen und deren Bearbeitung, informationelle Kosten, Arbeitszufriedenheit und andere Parameter als komplexe Bewertungsvariablen eingehen müßten.

2. Überblick über die Arbeitsziele des Vorhabens

Text-Parsing

TZ 1 Parsing von Textstrukturen

Der Verarbeitung von Volltexten zum Zwecke der Textkondensierung (automatisches Abstracting) wird beim Parsing-Mechanismus von TOPIC dadurch Rechnung getragen, daß eine gezielt textspezifische Systemlösung weiterentwickelt wird. Der unter besonderer Berücksichtigung von Texteigenschaften konzipierte Parser ist dadurch in der Lage, Textstrukturen zu berechnen. Die dabei auftretenden strukturellen Muster sind Ausdrucksmittel für textuelle Wohlgeformtheit und dienen dazu, die auf Phrasen- und Satzebene bestimmten Wissensstrukturen entsprechend ihrer Einbettung im Text repräsentationstechnisch zu überlagern, d.h. die Wissensbasis zusätzlich zu strukturieren. Entsprechendes Textwissen wird sowohl für die Textkondensierung (s. TZ 5) als auch für das Retrieval von Textinformation (s. TZ 7.1) benötigt.

TZ 1.1 Erkennung von Textkohäsion

Auf der lokalen Ebene der Textverkettung (Textkohäsion) wird die inhaltliche Expansion eines Textthemas und damit seine Abgrenzung gegenüber anderen benachbarten Textthemen festgelegt. Explizite sprachliche Ausdrucksmittel hierfür sind Konnektoren, implizite sind relationale Eigenschaften der im Weltwissen modellierten Konzepte. Damit werden in Texten z.B. zeitliche oder logische Abhängigkeiten zwischen Textthemen oder die Subordination eines Textthemas unter ein anderes angezeigt.

Die häufig intuitiven textlinguistischen Beschreibungen in das Format der in TOPIC entwickelten Wortexperten-Repräsentationssprache zu überführen, ist das eigentliche Forschungsproblem dieses Teilbereichs.

Zusätzlich werden im Rahmen dieses Teilziels auch Wortexperten-Lösungen für phrasen-orientierte sprachliche Phänomene (Negation u.ä.) entwickelt.

TZ 1.2 Erkennung von Textkohärenz

Auf der globaleren Ebene der Textwohlgeformtheit (Textkohärenz) werden Mechanismen behandelt, die mehrere Textthemen als Bestandteile einer inhaltlich zusammengehörigen textuellen Makrostruktur charakterisieren. Die Extensionsdaten der Textkohäsion werden damit funktional interpretiert, um unverbundene Textpassagen strukturell zu verknüpfen. Hierbei handelt es sich um bestimmte konventionalisierte Aufbau- und Abfolge-regularitäten in Texten, sog. Kohärenzrelationen, wie Typen von Spezialisierungs-/Generalisierungshierarchien von Aussagen, Stimmigkeit prozeduraler Abläufe usw., die den Text-Parsing-Prozeß durch textstrukturelle Erwartungen steuern und die Form der Wissensrepräsentation von Texten um textspezifische Konstrukte ergänzen.

Die dementsprechende Durchformung ergänzt die Textwissensbasis um eine Reihe von Indikatoren, die direkt für Kondensierungs- und Retrievalzwecke ausgenutzt werden. Eine Textzusammenfassung ohne Berücksichtigung dieser textpragmatischen Informationen würde damit auf einer entwickelteren Form des Faktenretrievals verbleiben. Dies erscheint für adäquate Textkondensate nicht hinreichend, wenn Textanalyse mehr ist als die kumulative Analyse von Einzelsätzen.

Im Unterschied zur Textkohäsion fallen bei der Untersuchung der Textkohärenz eigenständige Modellierungsprobleme an, die ganz besonders die Entwicklung retrievalunterstützender Kohärenzrelationen (etwa: Klassifikation, Vergleich usw.) betreffen.

Textrepräsentation

TZ 2 Speicherung mehrerer Textrepräsentationen in einer Wissensbasis

TZ 2.1 Entwicklung von inhaltsorientierten Auswahlmechanismen auf Textrepräsentationen

Werden mehrere Texte in einer Textwissensbasis repräsentiert, sind Operationen zur Auswahl einzelner Texte durch Spezifikation inhaltlicher Merkmale erforderlich. In einer Anfrage werden vorgegebene Repräsentationsmuster mit den Repräsentationsstrukturen der gespeicherten Texte verglichen und ein Text ausgewählt, wenn die Strukturen aufeinander passen. Aufgrund der stark vernetzten Wissensrepräsentationsstrukturen ergibt sich ein breites Spektrum an Möglichkeiten zur Festlegung, wann zwei Repräsentationsstrukturen aufeinander passen. Hier sind geeignete Verfahren zu entwickeln und zu implementieren.

TZ 2.2 Berücksichtigung von Änderungen der für die Analyse benutzten Weltwissensbasis

Im realen Einsatz wissensbasierter Informationssysteme ergeben sich Änderungen im zugrunde liegenden Weltwissen, z.B. durch Hinzunahme neuer Konzepte oder neuer Beschreibungsmerkmale existierender Konzepte. Solche Erweiterungen können Umstrukturierungen in der Wissensbasis zur Folge haben, die zu strukturellen Unterschieden zwischen den verschiedenen Versionen einer Wissensbasis führen. Aus verschiedenen Versionen erzeugte Textrepräsentationen sind dann ebenfalls in Teilbereichen inkompatibel. Damit kein Verlust relevanter Information in einer Retrievalanfrage entsteht, sind solche Fälle beim Entwurf und bei der Realisierung der Retrievaloperationen entsprechend zu berücksichtigen.

TZ 3 Wissensextraktion aus Texten

TZ 3.1 Erweiterung der verfügbaren Wissensrepräsentationsstrukturen

Aus den Ergebnissen der bisherigen Projektphase ergeben sich die Anforderungen an zwei neue Typen von Repräsentationsstrukturen. Die gleichzeitige Darstellung verschiedener Sichten auf ein Konzept sowie die Darstellung verschiedener Versionen eines Konzepts werden benötigt, um eine korrekte Zuordnung von Eigenschaften zu Konzepten vornehmen zu können und um von einem Basiskonzept abgeleitete Konzepte darzustellen. Aus der Erweiterung des Textparsers zur Erkennung von Kohärenzrelationen (vgl. TZ 1.2) ergeben sich weitere neue Typen von Repräsentationsstrukturen. Das Wissensrepräsentationsmodell und das Wissensverwaltungssystem sind in entsprechender Weise auszubauen.

TZ 3.2 Implementierung eines Interpreters für Integritätsregeln

Für manche Fälle von Wissensextraktion sind neben linguistischen Indikatoren Integritätsregeln erforderlich, die sowohl den Wissensextraktionsprozess selbst als auch die korrekte Einordnung des neuen Wissens in das bisher vorhandene unterstützen und möglich machen. Dazu sind die bisher realisierten, stark eingeschränkten Darstellungsmöglichkeiten für Integritätsregeln soweit auszubauen, daß aussagenlogische Terme formulierbar werden, und es ist ein Interpreter dafür zu implementieren.

Texttransformation

TZ 4-5 Verfahren zur Textkondensierung

Die Zusammenfassung der Volltexte in TOPIC wird primär wissensbasiert verlaufen, d.h. es dominieren die Mittel der Framerepräsentationsprache. Durch die Erweiterung der Text-Parsing-Komponente wird vermehrt auch textstrukturelles Wissen zur Textkohärenz bereitgestellt, das in die Textkondensierung einfließen soll. Zusätzlich sollen Erfahrungen aus der Informationspraxis umgesetzt und schon bekannte formale Typen der text-orientierten Informationsverwaltung in das Repertoire der Kondensat-Formate aufgenommen werden.

TZ 4 Erweiterung des Funktionsvorrats für die Textkondensierung

Über die auf einfache Interpretationen der Aktivierungsgewichte und die Zusammenhängigkeit einzelner Konzeptgruppen hinaus gehende Textkondensierungsfunktionen werden typische Struktureigenschaften von Frames (Slotbestand, Vernetzungs- und Dichtekriterien von Frame-Clustern usw.) einbeziehen. Die so berechneten Kondensate werden Themenschwerpunkte adäquater repräsentieren.

TZ 5 Integration textlinguistischen Kontrollwissens in die Berechnung von Textkondensaten

In die Textkondensierung werden neben Strukturdaten des Weltwissens (s. TZ 4) zunehmend auch textlinguistische Wohlgeformtheitskriterien eingehen. Diese stehen in direktem Zusammenhang mit den beim Text-Parsing identifizierten Textualitätsrelationen (thematische Elaboration, Vergleich, Klassifikation usw.) (s. TZ 1.2), müssen aber für die Zwecke der Textkondensierung ein Interpretationsschema durchlaufen, das Weltwissens- und Textstrukturdaten homogenisiert, d.h. unterschiedliche Hinweise auf Themenschwerpunkte eines Texts zur Erzeugung eines indikativen Kondensats vereinheitlichend zusammenfaßt. Dieses zu entwickelnde Interpretationsschema soll als Erzeugungsfiler für Kondensate wirken mit dem Ziel, die textuelle Stimmigkeit von Textkondensaten und damit die inhaltliche Kondensatakzeptanz für die Systembenutzer zu steigern.

Ergänzend hierzu sollen informationelle Filter treten, etwa mit der Vorgabe von Kondensat-Templates (Strukturmustern notwendiger inhaltlicher Aspekte von Kondensaten) oder SDI-Profilen, beides in der Informationspraxis häufig benutzte formale Typen für die Strukturierung von Informationsbedürfnissen (s. TZ 9). Diese Arbeiten sollten im Rahmen von Studienarbeiten vergeben werden.

Text-orientiertes Informationsmanagement

TZ 6 Graphisch-interaktive Komponente zur Navigation in Wissensnetzen

TZ 6.1 Entwicklung und Implementierung semantik-orientierter Layout-Strategien für Frame-Netze

Relationale Beziehungen zwischen Konzepten in semantischen Netzen werden bevorzugt mit Hilfe graphischer Darstellungen vermittelt. Bei steigender Zahl von Beziehungen besteht jedoch die Gefahr, daß die Darstellung unübersichtlich wird. Es sind Layout-Algorithmen zu entwickeln, die spezifische Struktureigenschaften des Netzes zur Gestaltung des Bildschirminhalts ausnutzen, um dem Benutzer Orientierungshilfen bei der Navigation zu bieten.

TZ 6.2 Sukzessives Expandieren von Teilnetzen

Der Umfang des Weltwissens wird es unmöglich machen, das Netz in seiner Gesamtheit am Bildschirm zu zeigen. Damit wird es notwendig, graphische Verfahren zu implementieren, die ein sukzessives Expandieren angedeuteter Teilnetze auf Bedarf ermöglichen, ohne den Benutzer durch weitgreifende Umordnungen des Bildschirminhalts zu irritieren.

TZ 7 Entwicklung einer graphischen Retrievalsprache

TZ 7.1 Graphische Präsentationsformen für Retrievalabfragen

Es sollen graphisch-interaktive Auswahlmöglichkeiten realisiert werden, die es dem Benutzer gestatten, Konzepte miteinander zu verknüpfen. Um den Reichtum der im Weltwissen modellierten Relationen auszunutzen, wird eine graphische Retrievalsprache entwickelt, die diesen Strukturen angepaßt ist.

TZ 7.2 Interpretation der im graphischen Dialog erstellten Abfrage als Suchauftrag an die Textwissensbasis

Aus den im Dialog erfaßten Spezifikationen der Suchbegriffe muß ein Repräsentationsmuster der Abfrage erstellt werden, das als Auftrag an die Textwissensverwaltung übermittelt wird. Dazu müssen Ausdrücke der in TZ 7.1 definierten Retrievalsprache, z.B. eingegrenzte Teilnetze aus dem Weltwissen, auf die Zugriffssprache des Textwissensverwaltungssystems abgebildet werden.

TZ 8 Anpassung der Dialogführung an die Retrievalsituation

Es sind Retrievalheuristiken zu entwickeln, welche Reaktionen des Systems in Problemsituationen, wie

- partielles matching
- zeitliche Variation des Weltwissens etc.

festlegen. Ziel der Heuristiken ist es, unter Einbeziehung und Interpretation der Dialoghistorie und der konkreten Konstellation der Textwissensbasis eine für den Benutzer informative Reaktion des Systems zu ermöglichen.

TZ 9 Faktenretrieval durch kaskadierten Zugriff auf Textwissen in abnehmenden Aggregierungsgraden

Graphisches Zooming und Expansion von Frames, Teilnetzen etc. sollen Zugang zur Textinformation in unterschiedlicher Detaillierung bieten. Dazu müssen Präsentationsoperationen implementiert werden, die je nach Abstraktionsniveau Textgraphen, Textkonstituenten, Frames, Slotseinträge und schließlich Textfragmente in einem einheitlichen Stil darstellen. Diese Ausgabefunktionen müssen durch eine Dialogführung unterstützt werden, die dem Benutzer jederzeit die Option einer Änderung des betrachteten Kondensierungsgrades offenhält.

3. Ausführliche Beschreibung der Arbeitsziele des Vorhabens

Textanalyse

Gegenstand der Textanalyse sind realistische Volltexte, und zwar nicht prä-editierte vollständige Zeitschriftenartikel (Umfang: ca. 2.000-4.000 Wörter) von einschlägigen deutschsprachigen Fachzeitschriften aus dem Bereich der Informations- und Kommunikationstechnik. Das Leistungsziel der Textanalyse in TOPIC ist auf die Erzeugung von Textzusammenfassungen (Kondensaten) aus diesen Volltexten auf einem indikativen Informationsniveau ausgerichtet.

Angesichts gewachsener Ansprüche an die Variabilität und Spezifität der Textanalyse werden die Anforderungen an das Text-Parsing und die Textrepräsentation sich erweitern, ohne bei gleichbleibendem Analyseziel, der groben semantischen Analyse von Volltexten (aboutness, shallow understanding), die Grundkonzepte des TOPIC-Systems zu tangieren:

- Parsing-Mechanismen, die aus heuristischen Überlegungen heraus eine eingeschränkte Erkennungskapazität besitzen und damit eine partielle Überführung der Texte in eine Repräsentationsstruktur leisten
- Wissensrepräsentationsformalismen, die (aus logischer Sicht) termbezogene Wissensrepräsentationsstrukturen (aus linguistischer Sicht: Nominalstrukturen) modellieren und damit ebenfalls eine partielle Repräsentation der Textinhalte unterstützen

Ausgangspunkt dieses Entwurfs für eine partielle, semantisch orientierte Textanalyse ist die Annahme, daß für das angestrebte Niveau der Textkondensierung eine voll ausdifferenzierte, umfassende Analyse der Volltexte nicht notwendig sei. Dies wird im Rahmen des partiellen Text-Parsing nunmehr in schlichter Form realisiert durch eine minimale Kollektion von in ihrer Erkennungsleistung eingeschränkten Wortexperten und eine auf Terme eingeschränkte Wissensrepräsentationssprache. Denn die Tiefe der Textanalyse ist bestimmt durch die Auswahl und Differenziertheit der verfügbaren Wortexperten und Weltwissenskonzepte. Auf den damit erzeugten Wissensrepräsentationsstrukturen operieren unterschiedliche Formen von Textkondensierungsfunktionen, um indikative Wissensstrukturen zu generieren.

Extensions- und Strukturdaten für Textthemen liefert der Text-Parser. Von eher untergeordneter Bedeutung ist für das Parsing die Erkennung von hauptsächlich (teil)satz-orientierten sprachlichen Phänomenen, wie Negation, Quantifikation usw. Allein zur Sicherung der erschöpfenden Darstellung von Informationen und Vermeidung unkorrekter Zustände des Weltwissens werden auch hierfür Wortexperten-Lösungen formuliert, ohne aber eine vollständige Behandlung der betroffenen lexikalischen Einheiten anzustreben. Zur Hypothese des Text-Parsing für die Zwecke der indikativen Textkondensierung gehört, daß phrasen- und satzorientierte Sprachphänomene bis zu einem gewissen Grad vernachlässigt werden können.

3.1 Text-Parsing

Der Verarbeitung von Volltexten zum Zwecke der Textkondensierung wird beim Parsing-Mechanismus von TOPIC dadurch Rechnung getragen, daß eine gezielt textspezifische Systemlösung ausgearbeitet wird. Dies entspricht dem Entwicklungsstand der Textlinguistik, in deren Forschungsbereich eine Vielfalt von sprachlichen Phänomenen entdeckt und strukturell beschrieben wurde, die die Annahme einer strukturellen Identität zwischen dem linguistischen Objekt "(schriftsprachlicher) Text" und dem "Satz" ausschließt. Es entspricht aber auch zugleich einem Defizit natürlichsprachlicher Systeme, wo die Verwendung dialog- orientierter Satz-Parser zur Entwicklung von meist ad hoc formulierten Hilfskonstrukten zur Erfassung von Dialogkohäsion und -kohärenz führt. Auf der Ebene der Systemperformanz läßt sich damit das Fehlen von Parsern, die auf Dialoggrammatiken beruhen, i.a. noch verdecken. Im Rahmen der Volltextverarbeitung verschärft sich dieses Modellierungsdefizit in einer Weise, die eine Verwendung text-orientierter Satz-Parser, also Parser ohne Bezug zu einer Textgrammatik, verbietet. Die Arbeiten im Bereich des Text- Parsing unterliegen dabei der Hypothese, daß die Leistungsmängel von solchen Parsern dann selbst auf die Ebene der Systemperformanz durchschlagen würden.

Der daraus für TOPIC abgeleitete Forschungsschwerpunkt "Text-Parsing" hat zum Gegenstand, die Konstruktion des Parsers direkt auf Textmerkmale abzustellen, um überhaupt Textstrukturen berechnen zu können. Die hier auftretenden strukturellen Muster sind Ausdrucksmittel für textuelle Wohlgeformtheit und dienen dazu, die auf Phrasen- und Satzebene bestimmten Wissensstrukturen entsprechend ihrer Einbettung im Text repräsentations-technisch zu überlagern, d.h. die Wissensbasis zusätzlich zu strukturieren.

TZ 1 Parsing von Textstrukturen

Beim Text-Parsing werden dazu zwei Formen von Organisationsmitteln für Textinhalte unterschieden: Textkohäsion legt die Expansion eines Textthemas und damit seine Abgrenzung gegenüber anderen Textthemen fest, Textkohärenz bewirkt die textuelle Relationierung mehrerer Textthemen. Entsprechende Informationen enthalten wesentliche Kondensierungs- und Retrievalhinweise; z.B. kann die Expansion eines mehrfach spezialisierten Textthemas allein aus strukturellen Gründen ab einer festzulegenden Explizitheitgrenze für das Kondensat abgebrochen werden, genauso wie ein kohärentes, aus dem Originaltext extrahiertes Textfragment allein einen hohen Relevanzwert für eine einschlägige Suchfrage hat.

TZ 1.1 Erkennung von Textkohäsion

Auf der lokalen Ebene der Textverkettung (Textkohäsion) sind bislang in TOPIC folgende Phänomene behandelt worden:

- (nominale) Anaphora
- lexikalische Kohäsion: Kollokation
- konjunktive / disjunktive Konnektoren

Aus den Leistungsmängeln des bislang realisierten Text-Parsers können nun konsequent noch fehlende Kohäsionskonstrukte bestimmt werden. Sie umfassen für den beantragten Projektzeitraum folgende, den Bereich der

Textkohäsion weitgehend erschöpfende Verkettungsmechanismen, die durch Konnektoren realisiert werden (Typologie nach LONGACRE):

- temporale Konnektoren
(z.B.: vor, nach, gleichzeitig, seitdem, unterdessen)
- teleologische Konnektoren
(z.B.: damit, um zu, weil, deswegen, daher)
- logische Konnektoren
(z.B.: obwohl, trotz, wenn-dann)
- konkatenative Konnektoren
(neben konjunktiven und disjunktiven) kontrastive Konnektoren
(z.B.: aber, dagegen)
- paraphrastische Konnektoren
(z.B.: mit anderen Worten, d.h.)
- exemplifizierende Konnektoren
(z.B.: z.B., etwa)

Ferner sind hier Kohäsionsphanomene zu erfassen, die über relationale Eigenschaften, die zwischen den in TOPIC modellierten Konzepten gelten, ausgedrückt werden, etwa die generalisierende oder exemplifizierende, parallele oder kontrastive Expansion eines Konzepts (etwa bei Rechner-Hardware vs. -Software).

Die häufig intuitiven textlinguistischen Beschreibungen in das Format der in TOPIC entwickelten Wortexperten-Repräsentationssprache zu überführen, ist das eigentliche Forschungsproblem dieses Teilbereichs.

TZ 1.2 Erkennung von Textkohärenz

Auf der Ebene der Textwohlgeformtheit (Textkohärenz) werden die lokal auftretenden Kohäsionsdaten nun funktional interpretiert im Sinne globaler Organisationsformen von textthematischen Makrostrukturen. Bereits behandelt ist in TOPIC die gesamte Typologie der Thema/Rhema-Relationen mit den beiden Grundmustern thematischer Progression, kaskadierendem Rhema und konstantem Thema, sowie mehreren Mischformen dieser Grundmuster. In Ergänzung dazu werden im beantragten Projektzeitraum weitere Strukturtypen der Textkohärenz modelliert und implementiert. Dazu zählen etwa, durchaus abgestimmt auf den behandelten Texttyp, Kohärenzrelationen des Typs:

- 1) Vergleich / Gegenüberstellung
Zwei oder mehr Objekte werden bezüglich einer Reihe von Merkmalen verglichen, wobei fortlaufend Vor- und Nachteile, günstige und ungünstige oder wünschbare und vorhandene Eigenschaften behandelt werden
- 2) Klassifikation
Objekte werden nach (Ad-hoc-)Klassen gruppiert und gemeinsame wie distinktive Merkmale innerhalb einer und zwischen den Klassen behandelt
- 3) unterschiedliche Zeitebenen
Objekte werden hinsichtlich unterscheidbarer Zeitperioden (etwa Rechnergenerationen, Marktphasen) bezüglich bekannter, hinzugekommener, noch fehlender usw. Eigenschaften beschrieben
- 4) prozedurale Stimmigkeit
Beschreibungs- und Handlungsabläufe unterliegen einer inhärenten auch in Texten reproduzierten Ordnung bei der Abfolge, Detaillierungstiefe usw.

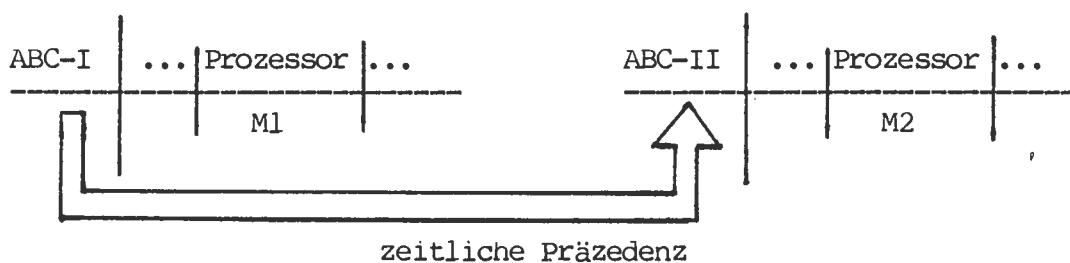
5) Text-Metastrukturen

Organisationsregeln für die Behandlung von Textinhalten auf meta-textueller Ebene (typische Organisationsmuster für Fachtexte, wie thematische Exposition-Methode-Ergebnis-Diskussion usf. oder rhetorische Figuren, wie These/Antithese, Exkurs usw.)

Im Unterschied zur Textkohäsion fallen bei der Untersuchung der Textkohärenz eigenständige Modellierungsprobleme an, die ganz besonders die Entwicklung retrievalunterstützender Kohärenzrelationen (etwa: Klassifikation, Vergleich usw.) betreffen.

Diese Textualität auf der Makroebene von Texten forciierenden Relationen ordnen sich in die folgende Skizze einer Texttheorie ein: Syntaktische Mechanismen auf der Phrasen- und Satzebene steuern die Interpretation lexikalisch kodierter Prädikate, wie sie in Form von Verben, Adjektiven oder Adverbien sprachlich realisiert werden. Die Argumentbelegung wird durch relativ statische Sortenspezifika kontrolliert, die einen grundlegenden informationellen Transfer zwischen Textproduzent und -rezipient garantieren. Die Zusammenhängigkeit eines Textthemas oberhalb der Phrasen- und Satzebene wird durch textuelle Kohäsionsmittel ausgedrückt. Die über diese Darstellungsmittel hinaus gehende dynamische Vernetzung zusammenhängender Themenkomplexe, die den "Informationswert" von Texten ebenso stark determiniert wie phrasen- oder satzinterne Strukturierungsmittel, wird über eine begrenzte Menge textstruktureller Muster gesteuert, den Kohärenzrelationen. Durchaus unabhängig vom in TOPIC gewählten Repräsentationsformalismus ist die prinzipielle Idee dabei, die als Träger dieser Kohärenzrelationen identifizierten Repräsentationseinheiten dementsprechend auch in der Wissensstruktur als kohärent zu repräsentieren. Für den Repräsentationsformalismus in TOPIC heißt das, die lokal durchaus schon stark vorstrukturierten Wissensseinheiten (Frames) durch Kohärenzrelationen dynamisch zu verknüpfen (und damit repräsentationstechnisch vorgegebene Strukturen textabhängig zu überlagern). Dies kann an folgendem einfachen Beispiel verdeutlicht werden (vgl. TZ 3.1):

Bsp: "... der Rechner ABC hatte ursprünglich einen M1-Prozessor. In seiner aktuellen Version dagegen verfügt er über einen M2-Prozessor ..."



Mit Hilfe dieser Überlagerungsmechanismen werden Kondensierungs- und Retrievaloperationen empirisch beschränkt (s. TZ 5 und TZ 7.1). Neben dieser systembezogenen Sicht sollte betont werden, daß damit textwissenschaftlich bislang eher diffuses Wissen über abstrakte Textorganisationsprinzipien mithilfe von Repräsentationskonstrukten operationalisiert und textgrammatisch systematisiert wird. Die zu Beginn postulierte Inadäquatheit von modifizierten Satz-Parsern für die Analyse von Volltexten, mithin die Notwendigkeit von Text-Parsern, manifestiert sich faktisch in der Erfassung von Textualitätsphänomen des oben beschriebenen Typs.

3.2 Textrepräsentation

In der ersten Projektphase von TOPIC wurde eine Wissensrepräsentationssprache entwickelt, die grundlegende, allen konzeptorientierten Wissensrepräsentationssprachen gemeinsame Repräsentationsstrukturen zur Verfügung stellt, wie z.B. Ober-/Unterbegriffshierarchie, Darstellung von Konzept-eigenschaften, Synonymie. In der zweiten Projektphase soll den speziellen Anforderungen, die sich aus der Repräsentation von Texten ergeben, Rechnung getragen und entsprechende Repräsentationsverfahren entwickelt werden.

TZ 2 Speicherung mehrerer Textrepräsentationen in einer Wissensbasis

In der bisherigen Projektlaufzeit lag der Arbeitsschwerpunkt für die Textrepräsentation auf der Entwicklung von Verfahren zur Repräsentation des Wissens eines analysierten Textes. In einem Informationssystem muß jedoch Wissen über mehrere Texte gespeichert und auf Anfrage zur Verfügung gestellt werden. Deshalb sollen in TOPIC II für die Probleme, die sich aus der Speicherung mehrerer Texte in einer Wissensbasis ergeben, Lösungen gesucht werden.

TZ 2.1 Entwicklung von inhaltsorientierten Auswahlmechanismen auf Textrepräsentationen

Werden mehrere Texte in einer Textwissensbasis repräsentiert, sind Operationen zur Auswahl einzelner Texte durch Spezifikation inhaltlicher Merkmale erforderlich. In einer Anfrage werden vorgegebene Repräsentationsmuster mit den Repräsentationsstrukturen der gespeicherten Texte verglichen und ein Text ausgewählt, wenn die Strukturen aufeinander passen. Aufgrund der stark vernetzten Wissensrepräsentationsstrukturen ergibt sich ein breites Spektrum an Möglichkeiten zur Festlegung, wann zwei Repräsentationsstrukturen aufeinander passen. Hier sind geeignete Verfahren zu entwickeln und zu implementieren.

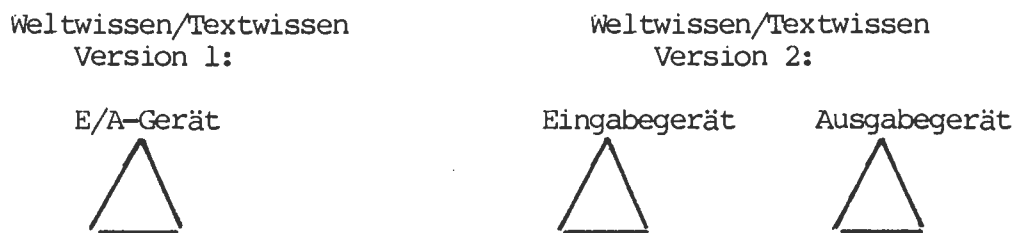
Konventionelles boolesches Retrieval, wo vorgegebene Zeichenketten mit der Menge von Zeichenketten, die einen Text kennzeichnen, verglichen werden, ist ein einfacher Spezialfall der zu entwickelnden Retrievaloperationen. Für sie müssen zusätzlich die relationalen Verkettungen zwischen Frames, die strukturelle Anreicherung von Frames durch Slots und Sloteträge, sowie die Beschreibungen der thematischen Struktur eines Textes als Ergebnis der Textkondensierungsprozeduren hinzu gezogen werden; in TOPIC II noch zu entwickelnde Erweiterungen der Repräsentationssprache (vgl. TZ 3.1) sind zusätzlich zu berücksichtigen.

Auf großen Textwissensbasen sind solche komplexen Retrievaloperationen nicht durch sequentielle Suche auf allen Textrepräsentationen durchführbar, sondern es muß ein globaler Repräsentationsteil vorgesehen werden, gegen den alle Suchfragen abgeglichen werden und der anschließend direkt zu allen relevanten Texten führt, also eine Art Indexstruktur für Wissensnetze ist. Hierzu sollen erste Überlegungen für eine Realisierung vorgenommen und eventuell auch Teile implementiert werden.

TZ 2.2 Berücksichtigung von Änderungen der für die Analyse benutzten Weltwissensbasis

Im realen Einsatz wissensbasierter Informationssysteme ergeben sich laufend Änderungen im zugrunde liegenden Weltwissen, z.B. durch Hinzunahme neuer Konzepte oder neuer Beschreibungsmerkmale zu existierenden Konzepten. Solche Erweiterungen können Umstrukturierungen in der Wissensbasis zur Folge haben, z.B. die Einführung neuer Oberbegriffe, Aufsplitten bisher nicht weiter differenzierter Begriffe in mehrere Unterbegriffe und daraus folgend wiederum Änderungen (non-terminaler) Slots anderer Frames. Die Folge sind strukturelle Unterschiede zwischen den verschiedenen Versionen der Weltwissensbasis. Aus verschiedenen Weltwissensversionen abgeleitete Textrepräsentationen sind dann ebenfalls in Teilbereichen inkompatibel. Die zu realisierenden Retrievaloperationen (vgl. TZ 7) haben diesen Punkt zu berücksichtigen, damit kein Verlust relevanter Information in einer Suchanfrage entsteht, weil die Repräsentationsstrukturen der Anfrage, die sich aus dem aktuellen Weltwissen ableiten, nicht mehr direkt auf die Repräsentationsstrukturen älterer Textwissensversionen passen. Neben einer entsprechenden Realisierung der Retrievaloperationen muß ein geeignetes Protokollierungsverfahren für Änderungen im Weltwissen entworfen und implementiert werden.

In folgendem einfachen Beispiel sind Ausschnitte von zwei Versionen des Weltwissens dargestellt:



In der ersten Version wird noch nicht unterschieden zwischen Eingabe- und Ausgabegerät. Eine Suchanfrage mit dem Zugriffskriterium "Eingabegerät" würde alle Texte, die mit dem alten Weltwissen analysiert wurden, unberücksichtigt lassen.

Eine erneute Analyse aller bisher bearbeiteten Texte nach jeder Änderung im Weltwissen wäre eine theoretische Lösung, ist aus Aufwandsgründen aber ausgeschlossen. Der einzig gangbare Weg besteht darin, alle Änderungen im Weltwissen fortlaufend zu protokollieren und diese Information in Suchfragen heranzuziehen, um auch die Textrepräsentationen, die auf alten Weltwissensversionen basieren, korrekt miteinzubeziehen. Die einzigen aus diesem Grund noch möglichen Ungenauigkeiten in einem Suchergebnis können sich daraus ergeben, daß das aktuelle Weltwissen spezifischer sein kann als alte Versionen. Im obigen Beispiel sind die Texte mit der alten Version nicht nach den Begriffen Eingabegerät und Ausgabegerät zu differenzieren, es kann lediglich durch Einbeziehung des Begriffs E/A-Gerät gewährleistet werden, daß relevante Texte nicht verloren gehen, aber nicht die Einbeziehung irrelevanter Texte in das Suchergebnis vermieden werden. In der Terminologie des klassischen Information Retrieval heißt dies, daß der Recall gesteigert wird, u.U. jedoch auf Kosten der Precision. Unterbegriffe von Eingabegerät oder Ausgabegerät, die im alten Weltwissen schon vorhanden waren, können aber durch das aktuelle, spezifischere Weltwissen im nachhinein beim Retrieval auch in alten Textrepräsentationen entsprechend klassifiziert werden. In diesem Fall erhöht sich der Recall, ohne die Precision zu verschlechtern.

TZ 3 Wissensextraktion aus Texten

In der ersten Projektphase wurden einige einfache Verfahren zur Extraktion neuen Wissens aus Texten entwickelt. So wird über die Abfrage von Slotfüllungsrestriktionen und unter Ausnutzung von Kontextbedingungen die Zuordnung von Eigenschaften zu Konzepten über Slotfüllung möglich. Die Aufnahme neuer, vorher unbekannter Konzepte in das Textwissen geschieht über die Erkennung von Eigennamen (im Fall von Produktnamen ist dies einfach) und über die Ausnutzung von Klassifikationshinweisen im unmittelbaren Kontext (z.B. "der Mikrocomputer TZ-X").

Die Erweiterung der Fähigkeit des TOPIC-Systems, neues Wissen aus Texten zu lernen, ist aus mehreren Gründen wünschenswert. Das Lernen neuer Sachverhalte, die zu strukturellen Umorganisationen in der Wissensbasis führen, dient der inkrementellen Erweiterung der Weltwissensbasis. Dies neue Wissen wird in periodischen Abständen von einem "knowledge engineer" kontrolliert in das Weltwissen aufgenommen (vgl. TZ 2.2). Darüber hinaus ist das Lernen neuen Wissens aus Texten aber für das eigentliche Textverstehen unmittelbare Notwendigkeit. Ein Text, der über einen Mikrocomputer TZ-X handelt, kann nur verstanden werden, wenn erkannt wird, daß der TZ-X ein (bisher unbekannter) Mikrocomputer ist. Die einfach zu fassenden Fälle sind, wie oben kurz erläutert, von TOPIC schon abgedeckt. Noch nicht erfaßte Fälle sind z.B. solche, wo im Text keine expliziten Hinweise zur kategoriellen Einordnung eines neuen Konzepts auftreten. Beispielsweise wurden in einem der bisher untersuchten Texte Programmpakete über ihre Namen eingeführt, ohne zu erwähnen, daß es sich dabei um Software handelt, diese Information war nur aus den Namen ableitbar: LisaCalc, LisaDraw, usf. Der Rest des Textes, der diese Programmpakete näher beschreibt, bleibt ohne die korrekte Einordnung dieser Begriffe weitgehend unverständlich, weil dort Konzepteigenschaften beschrieben werden, für die in der Textanalyse das Bezugskonzept fehlt. Eine Zuordnung dieser Eigenschaften kann somit nicht vorgenommen werden, und es gehen wichtige Fakten verloren.

TZ 3.1 Erweiterung der verfügbaren Wissensrepräsentationsstrukturen

Aus den Ergebnissen der bisherigen Projektphase ergeben sich die Anforderungen an zwei neue Typen von Repräsentationsstrukturen. Die gleichzeitige Darstellung verschiedener Sichten auf ein Konzept (z.B. ein Mikrocomputer als Rechner und als intelligentes Terminal) wird benötigt, um die Zuordnung von Einträgen zu den richtigen Slots vornehmen zu können, z.B. hat der Frame "intelligentes Terminal" andere Slots als der Frame "Mikrocomputer". Fehlt das Wissen über die möglichen Sichten auf ein Konzept, sind gewisse Slotfüllungen nicht durchführbar, und es gehen Fakten verloren. Es müssen Beziehungen zwischen verschiedenen Sichten eines Konzepts darstellbar sein, weil sie ebenfalls für eine korrekte und vollständige Repräsentation gebraucht werden (z.B. Überlappungen in den Slots, die dann gleiche Einträge besitzen müssen).

Die Darstellung verschiedener Versionen eines Konzepts wird erforderlich, um von einem Basiskonzept abgeleitete Konzepte (z.B. Ausbaustufen, Zukunftsversionen) darstellen zu können. Aus den gleichen Gründen wie für Sichten müssen zwischen Versionen des gleichen Konzepts Beziehungen modellierbar sein, beispielsweise können die Slotfüllungsbeschränkungen (z.B. für Speichergrößen) einer Zukunftsversion in bestimmter Weise weniger restriktiv sein als die des Basiskonzepts.

Die Hauptaufgabe in diesem Teilziel besteht darin, die Semantik dieser beiden Repräsentationsstrukturen herauszuarbeiten, die z.B. in Form von

Integritätsregeln (vgl. TZ 3.2) in das System eingebunden wird und damit eine korrekte Anwendung dieser Konstrukte gewährleistet. Die notwendigen Implementierungsmaßnahmen überlappen sich dabei recht stark, obwohl die Semantik von Sichten und Versionen unterschiedlich ist. Sich aus der Erweiterung des Textparsers noch zusätzlich ergebende und im Weltwissen zu erfassende Eigenschaften von Konzepten und ihrer Beziehungen untereinander (vgl. TZ 1.2) können ggf. durch Hinzunahme neuer Relationstypen dargestellt werden und erfordern deshalb nur einen geringen zusätzlichen Implementierungsaufwand. Auch für sie sind Regeln zur Semantikbeschreibung auszuarbeiten.

TZ 3.2 Implementierung eines Interpreters für Integritätsregeln

Für manche Fälle von Wissensextraktion sind neben linguistischen Indikatoren Integritätsregeln erforderlich, die sowohl den Wissensextraktionsprozeß selbst als auch die korrekte Einordnung des neuen Wissens in das bisher vorhandene unterstützen und möglich machen. Bisher darstellbare Integritätsregeln, die die Slotfüllung kontrollieren, beschränken sich auf die Angabe der Menge erlaubter Einträge. Vorstellbare Typen weitergehender Integritätsregeln spezifizieren z.B. Eigenschaften, die ein Frame besitzen muß, damit er ein zulässiger Slotseintrag ist (z.B. der Slot "Mikrocomputer" des Frames "CP/M 86" kann nur Mikrocomputer mit einer 8086-Cpu als Eintrag haben) oder beschreiben Beziehungen zwischen zwei Slots eines Frames (z.B. gibt es eine Abhängigkeit zwischen dem Slot "Preis" und dem Slot "Speichergröße" eines Plattenlaufwerks). Auch für neu hinzukommende Repräsentationskonstrukte (vgl. TZ 3.1) werden erweiterte Darstellungsmöglichkeiten für Integritätsregeln benötigt.

Die sich aus diesem Teilziel ergebenden Aufgaben bestehen in der Erweiterung der Wissensrepräsentationssprache zur Formulierung solcher Regeln sowie in der Implementierung eines Interpreters für aussagenlogische Ausdrücke. Die gleichen Erweiterungen werden für die Darstellung von Inferenzregeln notwendig, für die aber zusätzlich noch Ausführungszeitpunkte bzw. Triggerereignisse und Selektionskriterien angegeben werden müssen. Eine Berücksichtigung dieser Aspekte, die zur Implementierung einer Inferenzmaschine führt, kann voraussichtlich im Rahmen einer Diplomarbeit vorgenommen werden.

3.3 Texttransformation

Die Zusammenfassung (Kondensierung) der Volltexte in TOPIC wird primär wissensbasiert verlaufen, d.h. es dominieren die Mittel der Framerepräsentationssprache. Durch die Erweiterung der Text-Parsing-Komponente wird vermehrt auch textstrukturelles Wissen zur Textkohärenz bereitgestellt, das in die Textkondensierung einfließen soll. Zusätzlich wird die Option offen gehalten, Erfahrungen aus der Informationspraxis umzusetzen und schon bekannte formale Typen der text-orientierten Informationsverwaltung in das Repertoire der Kondensat-Formate aufzunehmen.

TZ 4 Erweiterung des Funktionsvorrats für die Textkondensierung

Die bislang dominierenden Zusammenfassungsmechanismen bei TOPIC beziehen sich auf einfache Interpretationen der Aktivierungsgewichte und die Zusammenhängigkeit einzelner Konzeptgruppen. In Erweiterung des schon entwickelten Konzepts des 'knowledge-based clustering' soll die Einbe-

ziehung von typischen Struktureigenschaften von Frames forciert werden, etwa in bezug auf die

- Füllungstiefe eines Slots
- Analogie von Aktivierungen von Slots mehrerer Frames
- Analogie von Füllungen von Slots mehrerer Frames
- Berücksichtigung von Wegelängen zwischen aktiven bzw. bewerteten ("gefüllten") Frames und Frame-Clustern
- Dichte von aktiven Frames in einem Cluster
- Interpretation von unterschiedlichen Populationstypen von Frame-Clustern
- Berechnung von Frame-Centroiden

Die Vielzahl dann verfügbarer Kondensierungsindikatoren verlangt eine weitere Differenzierung des bislang entwickelten Bewertungsmechanismus für thematische Schwerpunkte.

TZ 5 Integration textlinguistischen Kontrollwissens in die Berechnung von Textkondensaten

Ergebnis der bisherigen Projektarbeit im Bereich der Textkondensierung sind zwei noch völlig unverbundene Richtungen: Grundsätzlich realisiert ist die Berechnung von Textkondensaten ausschließlich aus Konfigurationsdaten des Weltwissens. Ebenfalls verfügbar ist die Berechnung von Grundtypen von Textkohärenzmustern, allerdings fehlt deren Einbindung in den Kontext der Textkondensierung. Diese beiden Stränge sollen jetzt verbunden werden derart, daß die Strukturierung von Texten nach textlinguistischen Wohlgeformtheitskriterien einfließt in die Berechnung von Textkondensaten.

Ausgangspunkt entsprechender Operationen werden die beim Text-Parsing identifizierten Textualitätsrelationen (thematische Elaboration, Vergleich, Klassifikation usw.) sein (s. TZ 1.2). Denkbare textlinguistische Kriterien für die Textkondensierung wären etwa der Abbruch von thematischen Expansionen eines Textthemas ab einer festzulegenden Explizitheitsstufe, eine Hierarchie für die Textkondensierung besonders relevanter Kohärenzrelationen o.ä. Die textlinguistischen Indikatoren müssen dann noch mit den aus dem Weltwissen abgeleiteten (s. TZ 4) nach zu entwickelnden Verträglichkeits- und Prioritätskriterien zu einem wissenbasierten, Textstrukturen integrierenden Kondensierungsverfahren abgestimmt werden.

Im Rahmen von Studienarbeiten wird die Option offen gehalten, bislang in der Informationspraxis erfolgreich benutzte referateähnliche Referenzmuster (Informationsdienstleistungen) für Dokumente auch in TOPIC II zu integrieren. Ergänzend zu textlinguistischen Einschränkungen der Textkondensierungsfunktionen könnten informationell tradierte Kondensatmuster eingesetzt werden, etwa mit der Vorgabe von Kondensat-Templates (Strukturmustern notwendiger inhaltlicher Aspekte von Kondensaten) oder SDI-Profilen, beides in der Informationspraxis häufig benutzte formale Typen für die Strukturierung von textuellen Informationsbedürfnissen.

3.4 Text-orientiertes Informationsmanagement

Hatte in der ersten Phase von TOPOGRAPHIC die Bereitstellung der Infrastruktur des Informationssystems (Rechnerkopplung) und die Entwicklung bzw. Anpassung graphischer Basissoftware das Ziel, zunächst die

Systementwicklung von TOPIC durch graphisch-interaktive Editoren zu unterstützen, so tritt in der zweiten Projektphase diese Form des Knowledge Engineering in den Hintergrund zu Gunsten der Implementierung einer an den Bedürfnissen des Endbenutzers ausgerichteten Retrievalschnittstelle. Die bei der Realisierung des Prototyps TOPOGRAPHIC-I gewonnenen Erfahrungen können in effizienter Weise genutzt werden, da die graphischen Softwareprimitive weitgehend in das inkrementelle Design des Interfaces integriert werden.

TZ 6 Graphisch-interaktive Komponente zur Navigation in Wissensnetzen

Wissensnetze zeichnen sich durch flexible Möglichkeiten zur Relationierung ihrer Bestandteile aus und ermöglichen eine effiziente Unterstützung bei der Suche nach relevanter Information, wobei die potentiell wichtigen Strukturinformationen gezielt genutzt werden können. Eine interaktive Navigation des Benutzers im Wissensnetz, basierend auf der Anwahl aktueller Konzepte in einer graphischen Präsentation des Netzes auf dem Bildschirm, wird ergänzt durch die wahlfreie Selektion von z.Zt. nicht sichtbaren Knoten durch Angabe ihrer Namens. Entscheidend für die Qualität der graphisch-interaktiven Navigation ist eine übersichtliche Darstellung des Netzes, wobei das Layout des aktuellen Teilnetzes Orientierungshilfen anbieten muß und die durch die Navigation bedingten Änderungen des Bildschirminhalts in überschaubarer Weise durchgeführt werden.

TZ 6.1 Entwicklung und Implementierung semantik-orientierter Layout-Strategien für Frame-Netze

Relationale Beziehungen zwischen Konzepten in semantischen Netzen werden bevorzugt mit Hilfe graphischer Darstellungen vermittelt. Bei steigender Zahl von Beziehungen besteht jedoch die Gefahr, daß die Darstellung unübersichtlich wird. Es sind Layout-Algorithmen zu entwickeln, die spezifische Struktureigenschaften des Netzes zur Gestaltung des Bildschirminhalts ausnutzen, um dem Benutzer Orientierungshilfen bei der Navigation zu bieten.

Die einzubeziehenden strukturellen Spezifika des Weltwissens ergeben sich bei der Analyse des Netzes unter verschiedenen Aspekten:

a) Formale Eigenschaften:

Sieht man das Wissensnetz als Zusammenfassung gerichteter Graphen, die jeweils eine Relation darstellen, so lassen sich mathematische Eigenschaften dieser Relationen aus ihrer formalen Definition im Kalkül des Frame-Datenmodells herleiten. Diese Eigenschaften können für das Layout wichtige Folgerungen implizieren:

- Symmetrie läßt auf eine starke inhaltliche Bindung der beteiligten Konzepte schließen, die deswegen auch in der Darstellung nicht zu weit voneinander entfernt sein sollten.
- Transitivität einer Relation ermöglicht das Darstellen eines sehr abstrakten Oberbegriffes und konkreter Unterbegriffe unter Auslassung der in der Hierarchie vorhandenen Zwischenstufen, falls die Konzepte mittleren Abstraktionsgrades für die Navigation nicht notwendig sind.

Es ist denkbar, daß auch andere mathematische Eigenschaften einer Relation Layoutentscheidungen beeinflussen können. Anhand der während des Projektverlaufs ins Welt- bzw. Textwissen aufgenommenen Relationen soll diese Frage beantwortet werden.

b) Inhaltliche Eigenschaften:

Auf einer pragmatischen Ebene ist das Netz nicht als abstrakte Struktur interessant, sondern die Beziehungen zwischen den repräsentierten Wissensinhalten treten in den Vordergrund. Die Modellierung semantischer Zusammenhänge der Begriffe resultiert in den Spezifika der Wissensrepräsentationssprache, insbesondere den Integritätsbedingungen, die für das Layout des Bildschirms genutzt werden. So weiß das System nicht nur, daß eine relationale Verkettung zweier Frames besteht, sondern kann auch den Grund dafür rekonstruieren, falls dies erforderlich wird. Die Is a-Relation zum Beispiel faßt verschiedene Arten der Konzeptspezialisierung zusammen, wie z.B.:

- Verfeinerung der Beschreibung durch Hinzunahme eines weiteren Merkmals (Slots).
- Extensionale Präzisierung des Konzepts durch Reduktion der Menge möglicher Einträge eines Slots.

Die Navigation kann nun effektiv unterstützt werden durch die Darstellung alternativer Klassen von Unterbegriffen eines Konzeptes, wobei eine Partitionierung nach der Spezialisierungsart vorliegt.

Die Größe und Komplexität des Wissensnetzes erfordern zu jedem Zeitpunkt des Dialogs die Festlegung des zu präsentierenden Teilnetzes. Bei der inhaltlichen Auswahl der abzubildenden Konzepte sind folgende Entscheidungen zu treffen:

- a) Welche und wieviele Relationen abgebildet werden, wird davon abhängen, ob die Navigation relations- oder konzeptorientiert erfolgt. Entweder spannen eine oder wenige Relationen ein Netz von zahlreichen Konzepten auf, oder wenige Konzepte werden in vollständiger Vernetzung gezeigt.
- b) Die Auswahl der Knoten resultiert in einer mehr lokalen oder einer regionalen Sicht auf das Netz, indem entweder Wert auf vollständige Darstellung benachbarter Knoten gelegt wird oder diese zugunsten entfernterer Konzepte übergangen werden können. In einer Hierarchie bzw. Multihierarchie hat dies Konsequenzen für die Zahl der "Generationen" und die Zahl der "Geschwister", die abgebildet werden können.

Welche Strategie zu bevorzugen ist, hängt von der jeweiligen Navigations- bzw. Retrievalsituation ab. Zur inhaltlichen Verfeinerung eines Konzepts während des Retrievals wird eine lokale konzeptorientierte Sicht von Vorteil sein, die inhaltliche Eigenschaften des modellierten Wissensfragments berücksichtigt. Ein regionaler relationsorientierter Netzausschnitt, der die formalen Charakteristika des Netzes (Hierarchie, Multihierarchie etc.) betont, kann eine bessere Übersicht bei der Navigation bieten.

Nachdem der Netzausschnitt ausgewählt ist, erfolgt das Bildschirmlayout unter folgenden Randbedingungen:

- a) Sowohl die Position der Knoten, wie auch die Linienführung der Kanten sind gemäß der Dialogsituation zu bestimmen. Dabei sind Verfahren der Graphentheorie, wie z. B. Planarisierung, weitgehend ungeeignet, da sie einerseits zu aufwendig sind, andererseits inhaltliche Strukturen des Weltwissens nicht berücksichtigen. Es werden daher heuristische

Verfahren entwickelt werden müssen, die sämtliche Informationen über das darzustellende Netz in einem integrierten Layout zusammenfassen.

- b) Infolge des selektiven Charakters der präsentierten Wissensausschnitte muß das Bildschirmlayout Information über zur Zeit nicht sichtbare Strukturen (nicht dargestellte Relationen, nur angedeutete Fortsetzungen des Netzes über den Rand des Bildschirms hinaus, etc.) als Orientierungshilfe beinhalten, um dem Benutzer alle Möglichkeiten zur Navigation offenzuhalten.

TZ 6.2 Sukzessives Expandieren von Teilnetzen

Der Umfang des Weltwissens wird es unmöglich machen, das Netz in seiner Gesamtheit zu zeigen. Damit wird es notwendig, graphische Verfahren zu implementieren, die ein sukzessives Expandieren angedeuteter Teilnetze nach Bedarf ermöglichen, ohne den Benutzer durch weitgreifende Umordnungen des Bildschirminhalts zu irritieren.

Da die Navigation - sieht man vom Sonderfall der direkten Eingabe des Zieles ab - durch Anwählen von Begriffen erfolgt, wobei die Folge ihrer Intensionen das gesuchte Konzept approximiert, befindet sich der Benutzer auf einer "Wanderung" durch das Wissensnetz. Um die jeweils aktuelle Umgebung des angewählten Knotens zu zeigen, muß sich der Bildschirmausschnitt dynamisch anpassen. Um konzeptuelle Sprünge zu vermeiden, darf das System den Bildschirminhalt nicht in einem Zug austauschen, sondern muß Zwischenschritte einführen, um eine inhaltliche Beziehung zwischen zwei aufeinanderfolgenden Dialogzuständen herstellen. So sollte zum Beispiel ein nicht mehr benötigter Teil des Netzes nicht schlagartig verschwinden oder überschrieben werden, sondern schrittweise aus dem Blickfeld wandern. Die für das statische Layout zu entwickelnden Strategien sind also um Verfahren zu ergänzen, die die Richtung eines Netzbildungsprozesses berücksichtigen.

Die Realisierung dieser Konzeption führt zu einer vom System gesteuerten Folge von graphischen Präsentationen ("dynamic graphics"), die den Eindruck einer "Animation" der Navigationsoperationen vermitteln. Dabei wird einer Implementierung der Vorzug gegeben werden, die das jeweils zu zeigende Teilnetz nach den Layoutstrategien auf dem Bildschirm rekonfiguriert, gegenüber der Präsentation des Gesamtnetzes als einer statischen Struktur, über die ein "Sichtfenster" bewegt wird. Ein dynamisch nach aktuellen Erfordernissen erstelltes Layout eines Teilnetzes hat folgende Vorteile:

- a) Angewählte Konzepte können automatisch so plaziert werden - durch sukzessive Bildschirmänderungen - , daß sie besondere Aufmerksamkeit erregen und damit die Orientierung erleichtern.
- b) Durch "Animation" von Netzteilen kann das Interesse des Benutzers auf Relationen und Konzepte gelenkt werden, die im Umfeld eines gerade neu gewählten aktuellen Begriffes angesiedelt sind. Diese Informationen ermöglichen eine zielgerichtete Navigation.

Neben der Sequentialisierung komplexer Systemreaktionen bietet die "Animation" die Möglichkeit, die Dialoghistorie (Trace) dem Benutzer zugänglich zu machen. Dies ist eine wichtige Orientierungshilfe, die nicht nur die Navigation, sondern auch die Konstruktion von Retrievalabfragen (vgl. TZ 7) wirkungsvoll unterstützt.

TZ 7 Entwicklung einer graphischen Retrievalsprache

TZ 7.1 Graphische Präsentationsformen für Retrievalabfragen

Der gezielte Zugriff auf die extrahierte Textinformation ist eine für die Systemperformanz essentielle Funktion, deren konzeptionelle Ausgestaltung die spezifischen Systemeigenschaften der Textwissensverwaltung in einen graphischen Dialog integriert. Es sollen graphisch-interaktive Auswahlmöglichkeiten realisiert werden, die es dem Benutzer gestatten, Konzepte miteinander zu verknüpfen. Um den Reichtum der im Weltwissen modellierten Relationen auszunutzen, wird eine graphische Retrievalsprache entwickelt, die diesen Strukturen angepaßt ist.

Prinzipiell wird die Formulierung einer Abfrage als Konstruktion eines Netzes realisiert, das die strukturelle und semantische Ausprägung der gesuchten Textinformation spezifiziert. Dem Benutzer werden zum sukzessiven Aufbau seines Abfragenetzes zwei Kommandotypen zur Verfügung gestellt:

a) Selektionsoperatoren

- Selektion und Freigabe von Konzepten

Das schon die Navigation prägende Anwählen von Konzepten mit der Maus stellt auch ein fundamentales Instrument zur Konstruktion von Abfragestrukturen aus den gezeigten Ausschnitten des Weltwissens dar. Natürlich benötigt man auch die Freigabe, falls ein Begriff nicht mehr in das aktuelle Interessenprofil paßt.

- Selektion und Freigabe von Relationen und Konzeptstrukturen

Das Anwählen von Relationen ist in verschiedenen Versionen zu implementieren:

- i) Selektion einer einzigen Kante (mit automatischer Anwahl der zugehörigen Knoten)
- ii) Selektion der auf den aktuellen Knoten gerichteten Kanten einer Relation
- iii) Selektion der von ihm ausgehenden Kanten
- iv) Selektion eines Teilnetzes (Teilbaumes)

Zusätzlich zu der relationalen Struktur, in die alle Konzepte auf der Netzebene integriert sind, werden zur Abfrageformulierung auch interne Strukturen der Frames bereitgestellt, so daß die Slots der Suchbegriffe zu Präzisierungen verwendet werden können.

Wie für die Konzeptselektion muß auch für diese Operationen die Möglichkeit bestehen, eine getroffene Entscheidung rückgängig zu machen oder teilweise - etwa bei gleichzeitiger Selektion aller auf ein Konzept gerichteter Kanten - zu revidieren.

b) Kombinationsoperatoren

Werden von den Selektionsoperatoren die Bestandteile des Navigationsnetzes zur Konstruktion der Abfragestruktur herangezogen, so bieten die Kombinationskommandos dem Benutzer die Möglichkeit, Beziehungen zu postulieren, die dann in den Textgraphen (teilweise) bestätigt werden, falls diese mit der Abfrage matchen. Zur Formulierung von Zusammenhängen, die nicht schon im Navigationsnetz modelliert sind, werden drei Kategorien von Relationen angeboten:

i) Logische Verknüpfungen

Die vom Booleschen Retrieval bekannten konjunktiven und disjunktiven Konnektoren und weitere logische Operatoren sollen implementiert werden.

Die in herkömmlichen Retrievalsystemen gebräuchlichen syntaktischen Kontextoperatoren werden durch leistungsfähigere, weil semantisch fundierte Operatoren ersetzt:

ii) Framebezogene Operatoren

Es soll die Möglichkeit bestehen, die Strukturen des Framedatenmodells zur Abfragekonstruktion zu nutzen, z.B. dem Slot PERIPHERIE des Frames MICROCOMPUTER eine Auswahl der Unterbegriffe des Frames PERIPHERIE zuzuordnen, der als Konzept an einer anderen Stelle im Navigationsnetz lokalisiert ist, um so die Menge der erlaubten Slotfüllungen zu spezifizieren. Die Konsistenzbedingungen des Weltwissens werden zur Kontrolle der konstruierten Abfragenetze herangezogen.

iii) Kohärenzrelationen

Die zur Repräsentation des Textwissens benutzten linguistischen Relationen bieten die Möglichkeit, die Abfrage nicht nur nach formalen und semantischen Kriterien zu formulieren, sondern auch die textuellen Eigenschaften der gesuchten Information anzugeben.

Die in herkömmlichen Retrievalsprachen gebräuchlichen Kommandos zur Outputsteuerung werden durch die in TZ 9 beschriebenen Möglichkeiten des interaktiven kaskadierten Abstracting ersetzt.

TZ 7.2 Interpretation der im graphischen Dialog erstellten Abfrage als Suchauftrag an die Textwissensbasis

Aus den im Dialog erfaßten Spezifikationen der Suchbegriffe muß ein Repräsentationsmuster der Abfrage erstellt werden, das als Auftrag an die Textwissenverwaltung übermittelt wird. Dazu müssen Ausdrücke der in TZ 7.1 definierten Retrievalsprache, z.B. eingegrenzte Teilnetze aus dem Weltwissen, auf die Zugriffssprache des Textwissensverwaltungssystems abgebildet werden.

Die Retrievalresultate werden von der Textwissensverwaltung mit fallendem Grad an Übereinstimmung mit der Abfragestruktur angeboten, so daß dem Benutzer zunächst das seinem Informationsbedürfnis am weitesten entgegenkommende Kondensat präsentiert werden kann. Er hat nun drei Möglichkeiten, den Retrievaldialog fortzusetzen:

- a) Er verfeinert sein Abfragenetz durch Integration von Teilen der gezeigten Textinformation oder durch Abänderung der Struktur nach erneuter Navigation (um das Netz durch weitere Konzepte zu ergänzen).
- b) Er wechselt zum nächsten der angebotenen Texte.
- c) Er entscheidet sich dafür, den Text in einem kaskadierten Abstracting (vgl. TZ 9) auf interessante Fakten hin zu untersuchen.

Die Implementierung muß diese potentiellen Fortsetzungen des Retrievaldialogs unterstützen und im Sinne eines zielgerichteten Retrieval zusätzlich eine Möglichkeit bieten, der Textwissensverwaltung den zu-

lässigen Grad der Abweichung eines Textgraphen von der Abfragestruktur als Parameter zu übergeben.

TZ 8 Anpassung der Dialogführung an die Retrievalsituation

Der in den Teilzielen 6 und 7 in seiner Funktionalität spezifizierte Mensch-Maschine-Dialog ist gekennzeichnet durch ein hohes Maß an Variabilität und Komplexität. Das skizzierte Design der Systemoberfläche versucht, durch den Einsatz graphischer Interaktionsformen, wie z.B. Selektion von Objekten, den Endbenutzer von Anforderungen syntaktischer Art (korrekte Eingabe von Kommandos etc.) weitgehend zu befreien. Angesichts der komplexen Strukturen - charakteristisch für die zu präsentierende Information - und der nicht auf ein starres Handlungsschema reduzierbaren Möglichkeiten der Dialogführung wird das Interface einen hohen Flexibilitätsgrad in seinen Reaktionen aufweisen müssen.

Insbesondere in "Konfliktsituationen", in denen dem Wunsch des Benutzers nicht vollständig entsprochen werden kann oder inkonsistente Informationen vorliegen, sollte der Benutzer durch eine der Situation angepaßte Reaktion des Systems bei seinen Entscheidungen unterstützt werden. Probleme dieser Art sind z.B.:

- "partiell matching":

Kann die Abfragestruktur nur partiell von einer (oder mehreren) der gespeicherten Textrepräsentationen abgedeckt werden, so muß die Relevanz des Textes situationsspezifisch bestimmt werden.

- zeitliche Variation des Weltwissens:

Wie in TZ 2.2 erläutert, wird der Fall auftreten, daß den gefundenen Textrepräsentationen unterschiedliche konzeptuelle Strukturen zugrunde liegen. In Abhängigkeit von dem bislang geführten Dialog wird auch hier die Relevanz der Texte zu ermitteln sein.

Vor diesem Hintergrund erscheint im Sinne einer Performanzsicherung des Gesamtsystems die Zielsetzung plausibel, durch die Einbeziehung verschiedener Wissensquellen ein möglichst fundiertes und damit intelligentes Systemverhalten zu erreichen. Die relevanten Informationen zeichnen sich durch unterschiedliche Grade von Situationsgebundenheit aus, beginnend mit dialogsyntaktisch orientierten Zustandsdaten (Füllungsgrad des Bildschirms etc.) bis zu den spezifischen Welt- und Textwissensstrukturen, die für ein adaptives Verhalten benötigt werden.

Der effiziente Einsatz der Möglichkeiten der Systemoberfläche erfordert die koordinierte Interpretation von Informationen aus den folgenden Wissensquellen:

a) Dialoggedächtnis und -heuristik:

Eine Klassifizierung der zu implementierenden Operationen ermöglicht die Modellierung der Dialogschritte, die entscheidungsrelevante Information zur Determination der Systemreaktion bereitstellen soll. So wird das Layout mehr Übersichtsinformation bieten, wenn der bisherige Dialogverlauf auf ein "Browsing" des Benutzers schließen läßt, dagegen aber Detailinformation anbieten, wenn die Navigation zielgerichtet (etwa eine Hierarchie hinabsteigend) erfolgt.

- b) Graphische Daten der Objekte:
Die Information über das vorliegende Bildschirmlayout (Füllungsgrad, relations- bzw. konzeptorientierte Darstellung (vgl. TZ 6.1), etc.) kann die Dialogführung im Sinne einer Harmonisierung der Änderungen auf der Systemoberfläche beeinflussen.
- c) Welt- bzw. Textwissen:
Die systeminhärente Strukturinformation des Weltwissens könnte z.B. genutzt werden, um dem Benutzer vorzuschlagen, Oberbegriffe von Konzepten aus der Abfragestruktur zu entfernen und dies mit seinem Einverständnis automatisch durchzuführen, wenn zuviele Textgraphen partiell mit der Abfragestruktur übereinstimmen.

Die flexiblen Reaktionsmöglichkeiten des Interface sollen auf heuristischen Navigations- und Retrievalregeln beruhen (s. obige Beispiele), deren Plausibilität zwar anhand der Literatur überprüft wird, jedoch im Rahmen des beantragten Projekts nicht empirisch begründet werden soll.

TZ 9 Faktenretrieval durch kaskadierten Zugriff auf Textwissen in abnehmenden Aggregierungsgraden

Graphisches Zooming und Expansion von Frames, Teilnetzen etc. sollen Zugang zur Textinformation in unterschiedlicher Detaillierung bieten. Dazu müssen Präsentationsoperationen implementiert werden, die - je nach Abstraktionsniveau - Textgraphen, Textkonstituenten, Frames, SlotEinträge und schließlich Textfragmente in einem einheitlichen Stil darstellen. Diese Ausgabefunktionen müssen durch eine Dialogführung unterstützt werden, die dem Benutzer jederzeit die Option einer Änderung des betrachteten Kondensierungsgrades offenhält.

Für das Interface stellt sich damit das Problem, jede Kondensierungsstufe in einer ihrer Struktur angemessenen Art darzustellen und zugänglich zu machen. Unter diesem Gesichtspunkt gilt es, die bisher spezifizierten Präsentationsverfahren und Kommandostrukturen an die erweiterten Aufgaben anzupassen.

a) Präsentation von Kondensaten

Der konzeptuelle Abstand zwischen den verschiedenen Informationsniveaus spiegelt sich in einer sehr inhomogenen Struktur, die von Frames und Netzen bis zu Volltexten reicht. Ein Darstellungsverfahren, das auf eine Datenstruktur hin zugeschnitten ist, würde diesem Umstand nicht gerecht werden. Es könnte graphisch kaum befriedigend realisiert werden und würde außerdem eher dazu beitragen, inhaltliche Unterschiede zu verwischen. Auf die einzelnen Abstraktionsstufen zugeschnittene Einzellösungen würden den Eindruck einer homogenen Systemfunktionalität zerstören. Die strukturelle Anpassung der Darstellung muß deshalb einer einheitlichen Dialogstrategie untergeordnet werden. Ebenso wie auf inhaltlicher Ebene der Abstand zwischen einer quasi stichwortartigen Inhaltsangabe und dem Volltext durch kleine Schritte überwunden wird, soll sich auch die Darstellung von Volltexten aus den Präsentationsformen für Netze entwickeln. Als Bindeglied werden Tabellen verwendet, die in netzwerkorientierten Abbildungen Detailinformationen über einzelne Knoten enthalten. Indem die Textsicht weniger global wird, verliert die Strukturinformation an Bedeutung gegenüber der Fakteninformation, die rein tabellarisch vermittelt wird, wobei als Einträge zunächst Elemente des Textgraphen schließlich aber auch Fragmente des Volltextes auftauchen. Als letzte Stufe der Kaskade wird der Originaltext aufgelistet, wobei die Extrakte zunächst ganze Sätze, dann Absätze und schließlich den gesamten Text umfassen.

b) Navigation im Textwissen.

Damit das Textwissen dem Benutzer zugänglich wird, muß er in die Lage versetzt werden, sich in den Wissensstrukturen zu bewegen. Dabei lassen sich folgende Bewegungsrichtungen unterscheiden:

- Wechseln des Textes:

Nach Bearbeitung der Suchanfrage stellt die Textwissensverwaltung dem Benutzer einen exemplarischen Text zur Verfügung, den er zunächst betrachten kann. Entweder wird er sich dann weitere aufgefundene Texte zeigen lassen oder seine Anfrage mit Hilfe von Textinformationen verfeinern (retrieval by example).

- Bewegung entlang den Stufen des kaskadierten Abstracting:

Der Benutzer will den Text unter einer verfeinerten (vergrößerten) Sicht untersuchen. Er kann dabei der vorgegebenen Stufung folgen oder auch einzelne Schritte überspringen.

- Navigation innerhalb eines Kondensats, Textgraphen etc.:

Sind Text und Abstraktionsniveau festgelegt, kann die somit angewählte Wissensstruktur nach Informationen durchsucht werden. Dabei werden Navigationsoperatoren zum Durchwandern von Relationen in Textgraphen und zum Durchblättern von Texten nötig.

Um eine homogene Kommandostruktur des Gesamtsystems zu erreichen, sollten analoge Operatoren modusunabhängig von gleichen Kommandos realisiert werden. Es sollte also versucht werden, die bisher schon für Navigations- und Retrieval definierten Kommandos für die Textwissensnavigation einzusetzen und so zu einem kleinen überschaubaren Befehlssatz zu kommen. Die Bedeutung von Operatoren wie "Select" oder "Scroll" zum Beispiel ist so gut zu verallgemeinern, daß sie sowohl auf Texte wie auch auf Netze anzuwenden sind.

Folgerungen für die Systemstruktur des Interface

Die interaktiven Features der Systemoberfläche konstituieren eine graphische Kommandosprache, die einen integrierten Zugang zu den Systemfunktionen Information Management bzw. Retrieval ermöglicht. Gegenüber den in der Informationspraxis verbreiteten Systemschnittstellen, die häufig auf eine einzige, stark vorstrukturierte Aufgabe und eine bestimmte, von der Systemanalyse determinierte Benutzerklasse fixiert sind, erwachsen bei dem projektierten multifunktionalen Interface erheblich größere Anforderungen an die Flexibilität des Systemverhaltens. Während die heterogene Funktionalität die Einführung unterschiedlicher Arbeitsmodi - zugeschnitten auf Aufgabe und Situation - bedingt, wird das Problem der benutzeradäquaten Systemreaktionen durch den speziellen Charakter der zu implementierenden Funktionen (Navigation, Retrieval, Kaskadiertes Abstracting) verschärft. Einerseits handelt es sich um komplexe, nicht vollständig strukturierbare Tätigkeiten, andererseits muß das Verhalten des Systems kontextabhängig (Dialogsituation, Charakteristika der zu präsentierenden Informationen) konzipiert sein, um eine sinnvolle Unterstützung des jeweiligen Benutzers zu gewährleisten.

Herkömmliche Techniken der Dialogsteuerung (endliche Automaten, Netze) erlauben nur einen vom Designer vorstrukturierten Dialog oder die Auswahl aus einigen Dialogvarianten. Die projektierte Schnittstelle muß demnach alternative Repräsentationsformen einsetzen, um den Anforderungen an die Flexibilität gerecht zu werden. Erschwerend wirkt sich die Inhomogenität

der zu speichernden Informationen aus, die bewirkt, daß allzu sehr an die Struktur der jeweiligen Daten angepaßte Repräsentationen nicht verwendet werden können, da sie die koordinierte Interpretation der aus den verschiedenen Wissensquellen stammenden Informationen verhindern.

Die Steuerung des Dialogs erfolgt durch Inferenzen, die auf den vorhandenen Wissensbasen durchgeführt werden. Dazu sind die in den Teilzielen 6 bis 9 z.T. bereits skizzierten Heuristiken als Regeln der Form:

⟨ Situation ⟩ → ⟨ Aktion ⟩

zu formulieren. Zur Durchführung der Inferenzen benötigt man entsprechende Software, deren Entwicklung ein technisches Teilziel des Projektes darstellt.

4. State-of-the-Art

4.1 Übersicht über Methoden der Volltextverarbeitung in Informationssystemen

Im folgenden werden fünf methodische Ansätze unterschieden, nach denen Volltexte bislang im Kontext von Informationssystemen repräsentiert und auf diesen Repräsentationsstrukturen von Volltexten Operationen ausgeführt werden, um informationell relevante Text(kondensat)e zu ermitteln (erzeugen) und in für den Benutzer adäquater Darstellung bereitzustellen. Denn anders als die Verarbeitung von Textsubstraten (Referate, Titel oder koordinierte Deskriptoren) in konventionellen Dokument-Retrieval-Systemen verlangen Volltexte (vollständige Zeitschriftenaufsätze, Briefe, Protokolle, Gesetzestexte, Zeitungsartikel oder gar Bücher) mindestens zusätzliche - wenn nicht qualitativ neue - Aufbereitungs- und Wiedergewinnungsverfahren und damit Methoden der informationellen Aggregation (Textkondensierung) sowie Verwaltung und Präsentation (Textinformationsmanagement). Es sollen zunächst nur die grundlegenden Prinzipien der verschiedenen Verfahrenstypen - des informatischen, statistischen, linguistischen, wissensbasierten und kognitiven Ansatzes - für die Volltextverarbeitung grob skizziert werden. Dabei erfolgt die Sicht auf diese Verfahrensgruppen aus dem Blickwinkel der informationellen Textanalyse und der Volltexten adäquaten Verwaltungs- und Präsentationsformen. Soweit die im Rahmen dieser Ansätze entwickelten Konzeptionen direkt für spezielle Teilbereiche von TOPIC oder TOPOGRAPHIC einschlägig sind, werden sie in den entsprechenden Kapiteln dieses Literaturberichts weiter ausgeführt (Kap. 4.2). Unter Volltextverarbeitung wird im Kontext der Informationswissenschaft - im Unterschied zum eingeschränkteren Verständnis des Begriffs in der Informatik (zur Unterscheidung vgl. SALLIS 1978a) - sowohl die Überführung eines Eingabetexts in eine (semantische) Repräsentationsform als auch Operationen darauf sowie die Erzeugung/Bereitstellung eines Ausgabetexts verstanden. Der folgende Bericht konzentriert sich im wesentlichen auf die für das beantragte Projekt zentrale transformationelle Grundoperationen auf Textrepräsentationsstrukturen, die Kondensierung, und mit ihr in Beziehung stehende Varianten des Textinformationsmanagements. Die Kondensierung ist eine bedeutungsreduzierende Transformation von Textinhalten (im Unterschied etwa zur bedeutungserhaltenden Paraphrase). Textkondensate können sein: Schlagwörter (Indexing) oder Klassifikatoren (Klassifikation) für Texte, indikative oder informative Zusammenfassungen (Referate) von Texten (Abstracting). Grenzfälle, wie das Retrieval einzelner Fakten oder relevanter Extrakte, d.h. originaler Passagen aus einem Volltext, werden ebenfalls mit einbezogen.

4.1.1 Der informatische Ansatz

Unter dem informatischen Ansatz werden all jene Repräsentations- und Transformationskonzepte gefaßt, die keine Verfahren der eigentlichen Informationsaufbereitung ("Informationsveredelung") für Volltexte bereitstellen - sieht man von eher trivialen Fällen der Stoppworteliminierung ab -, sondern lediglich eine Informationsbereitstellungsfunktion haben. Traditionell im Mittelpunkt stehen dabei auf Platz- und Zugriffsoptimierung ausgelegte Datenhaltungs- und Speicherorganisationsformen für Textdateien (inverted file, hash-Transformationen, string compression etc. (RABITTI/ZIZKA 1984)), aber auch speziell auf das Text-Retrieval abgestimmte Hardwarelösungen für Such- und Speichertechnologien (HOLLAAR 1983). Aktuelle Entwicklungen, die auf eine Integration von

Datenbankmethodik und Text-Informationssystemen abzielen, modellieren lediglich den Aspekt der Daten-/Textverwaltung (SCHEK 1981), ohne damit das Textanalyse-Defizit des informatischen Ansatzes zu vermindern, d.h. Repräsentationsstrukturen für den Inhalt von Texten bereitzustellen, die über die stoppwortbereinigte Tokenkollektion des Ausgangstexts hinausgehen.

Informationswiedergewinnung auf dem Niveau des "post co-ordinate indexing" charakterisiert die Funktionalität dieser Systeme für das Textretrieval. Einfache Suchverfahren in formatierten Textdateien werden in kommerziellen Systemen über Anfragesprachen mit Booleschen und/oder Kontext-Operatoren gesteuert. Als Folge des oben angesprochenen Integrationstrends werden in experimentellen Systemen Anfragesprachen von Text-Informationssystemen um relationenalgebraische Operatoren (MACLEOD 1981) erweitert. Unverändert bleibt dabei die für informatische Systeme charakteristische Retrieval-Philosophie des einfachen String-Matching zwischen Suchfrage und Termbestand der gesamten Dokumentenkollektion, die durch Optionen für systemgestützte Frageerweiterungen (down-posting/up-posting) aufgrund semantischer Vorgaben eines Hintergrundwörterbuchs (Thesaurus) nur wenig gemildert wird. Es werden dann alle Texte bzw. einschlägigen Textpassagen (O'CONNOR 1973) ermittelt und bereitgestellt, deren Termbestand mit der Spezifikation der Suchfrage exakt übereinstimmt.

Im Zuge des "traditionellen" Ansatzes (vgl. INGWERSEN 1984), die offensichtlichen Performanzdefizite bestehender Retrievalsysteme durch Erweiterungen des angebotenen Funktionsspektrums (z.B. Freitextoperatoren, "crossfile search") abzumildern, werden von einigen Retrievalinterfaces Features angeboten, die den Dialog zwar komfortabler gestalten, aber die eigentlichen, d.h. konzeptionellen Schwachpunkte des Systems nicht tangieren. So bietet das CALIBAN-System (FREI/JAUSLIN 1983) die Möglichkeit, in einer graphischen Präsentation des INSPEC-Thesaurus interaktiv zu navigieren, nutzt aber die gespeicherten semantischen Informationen nicht zur Steuerung der Dialogführung. Andere Systeme bieten kommunikationstechnische Serviceleistungen (automatisches Anwählen von Hosts etc.), die auf "intelligenten Terminals" (Intelligenz im Sinne der Verfügbarkeit lokaler Rechenkapazität) basieren (vgl. TOLIVER 1982).

Der informatische Ansatz wird gerade in der Informationspraxis durch den Aufbau sog. Volltext-Informationbanken (für Zeitungen, Gesetzestexte, Parlamentsprotokolle usf.) momentan mit Vehemenz verfolgt. Die Überlegung, Volltextverarbeitung aber nicht nur als im oben skizzierten Sinne "technologisches", sondern primär als ein informationelles Repräsentations- und Transformationsproblem aufzufassen, wie es der vorliegende Projektantrag nahelegt, gewinnt durch offensichtliche Probleme mit Volltext-Informationssystemen (MARON 1982, HAHN 1983) zunehmend an Plausibilität. Zu den einschneidendsten Mängeln zählen: die formalisierte Mensch-Maschine Interaktion (z.T. bei grafik-gestütztem Information Retrieval aufgehoben), die grobe Vereinfachung syntaktischer und semantischer Aspekte von Texten, die zur benutzerseitigen Antizipation sprachlicher Regularitäten (etwa beim Gebrauch der Kontext-Operatoren) zwingt oder gar aus systematischen Gründen auch benutzerseitig nicht mehr aufgefangen werden kann (bedingt etwa durch Restriktionen der Anfragesprachen, die die Variabilität natürlicher Sprachen nicht aufzufangen vermögen). Das oben angesprochene Manko informatischer Systeme, Texte lediglich bereitzustellen, statt sie analytisch aufzubereiten, zeigt sich an der einfachen Referenzleistung, aufgrund eines spezifizierten Suchmusters ohne texttransformationelle Zwischenstufen einen Originaltext verfügbar zu machen.

4.1.2 Der statistische Ansatz

Der statistische Ansatz zur Volltextverarbeitung basiert auf der Berechnung der Zusammenhängigkeit (Ähnlichkeit) oder Signifikanz (Wichtigkeit) von Terms aufgrund von Ko-Okkurrenzen oder Häufigkeitsverteilungen in Dokument(kollektion)en. Interpretiert man entsprechende statistische Maßzahlen als semantische Indikatoren, so werden dadurch konzeptuelle Bezüge auf der lexikalischen Ebene im Dokumentbestand - sei es durch die Berechnung von charakterisierenden Index-Terms für ein Dokument (Indexing: DILLON/McDONALD 1983) oder die Gruppierung mehrerer Dokumente auf der Basis ähnlicher Termbestände (Klassifikation: BOREHAM/NIBLETT 1976) - explizit gemacht. Dieser insgesamt erfolgreich auf das Information Retrieval von Abstracts (RIJSBERGEN 1979) angewandte Ansatz bringt jedoch für die Volltextverarbeitung erhebliche Komplexitätsbarrieren mit sich (z.B. die Multiplikation großer Matrizen bei der Berechnung von Assoziationsfaktoren), die durch heuristische Mittel unterlaufen werden müssen (z.B. durch lokales Clustering auf der Grundlage von Feedback-Daten (ATTAR/FRAENKEL 1977)). Die systematische Beschränkung des statistischen Ansatzes, informationelle Bezüge ausschließlich zwischen Schlagwörtern und Volltexten herzustellen, wird nur scheinbar durchbrochen beim statistischen Abstracting. Ausgehend von Termhäufigkeiten werden diejenigen Originalsätze eines Texts als besonders relevant ausgewählt, in denen mehrere statistisch signifikante Einzelterms zugleich auftreten, und in Form eines Text-Extrakts, d.i. die Menge aller über dieses Signifikanzkriterium ausgewählten Einzelsätze des Originaltexts, wieder zusammengesetzt (Überblicksdarstellung zu diesem Ansatz: WYLLYS 1967). Statistische Signifikanz von Terms kann durch zusätzliche Indikatoren (a priori gesetzte Schlüsselwörter, gleichzeitiges Auftreten des Terms als Titelwort, signifikante Position im Text (Textanfang/-ende), relative Häufigkeit von substantivischen Neologismen) angereichert werden (EDMUNDSON 1969, KARASEV 1974). Eine weitergehende Ausdehnung des statistischen Ansatzes auf die Textebene - im Sinne eines nicht als Extracting verstandenen statistisch orientierten Abstracting-Ansatzes - aber ist prinzipiell (wegen der Äquivalenz probabilistischer Sprachmodelle mit regulären Sprachen/endlichen Automaten und der Unbrauchbarkeit dieser Konzepte zur Beschreibung natürlicher Sprachen) ausgeschlossen. Der statistische Ansatz zur Volltextverarbeitung erscheint systematisch auf lexikalische Transformationsleistungen (Indexing, Klassifikation) beschränkt.

Statistische Konzepte sind auch in Formen der Dateiorganisation (clustered files), Volltext-Suchheuristiken (MAH/D'AMORE 1983) und das Textretrieval (CROFT/ROCHFORT 1981) eingegangen, etwa bei der Suchformulierung durch Gewichtung der Terms in der Suchfrage und, entsprechend dem Grad der Übereinstimmung zwischen gewichteter Suchfrage und Dokumentbeschreibung, der Ausgabe nach Rangfolge der Dokumente (ranking). Am Prinzip quasi-Boolescher Anfragesprachen ändern diese Erweiterungen aber nichts. Genauso wie die Kondensierung des Texts im allgemeinen die globale Referenzleistung auf ein Original-Dokument unterstützt (vgl. aber auch die Berechnung signifikanter Paragraphen eines Original-Texts in der SMART-Extension ANNOD (BERNSTEIN/WILLIAMSON 1984)).

4.1.3 Der linguistische Ansatz

Linguistische Ansätze zur Repräsentation und Transformation von Volltexten sind grundsätzlich regelbasiert und damit oberhalb des einfachen Pattern Matching von Texttoken mit Positiv- oder Negativlisten (vgl. dazu ARTANDI 1969) angesiedelt.

Eher informationswissenschaftlicher Tradition gehorchend als dem Potential dieses Ansatzes gerecht zu werden, ist in unmittelbarer Konkurrenz zu statistischen Ansätzen eine linguistische Variante für das Volltext-Indexing entwickelt worden. Dabei wurde dem pragmatischen Konzept des partiellen Parsing von Nominalphrasen (ROSTEK 1979) - einem echten informationslinguistischen Methodenbeitrag zur maschinellen Sprachverarbeitung - die linguistisch "saubere" (aber für das Ziel der Deskriptorenerkennung vermutlich unnötig aufwendige) Lösung des vollständigen Parsing als Methodenalternative gegenübergestellt (etwa durch den SATAN-Parser im Systemkomplex JUDO (FELZMANN et al. 1980)).

Linguistische Verfahren, und das heißt in Verschärfung von regelbasierten Verfahren hier die Einbeziehung von Grammatiken, beziehen ihre unbestreitbare Legitimation jedoch aus qualitativ anderen Sektoren der Volltextverarbeitung, etwa der Überführung unformatierter Texte in formatierte Daten einer Datenbank (SAGER 1978) bzw. der Generierung von Texten aus solchen formatierten Daten. Letztere, vorwiegend dem Bereich der medizinischen Informationslinguistik zuzuordnende Versuche, Fließtexte aus formatierten Daten zu erzeugen, sind bislang häufig ohne besondere linguistische Systematik entwickelt worden (MARTIN et al. 1977) und eher ein Spezialfall des formatierten Datenretrieval, das konventionell als reine Extraktion relevanter Fakten (HIRSCHMAN/GRISHMAN 1977), aber auch mit dem Komfort natürlichsprachlicher Frage/Antwort-Umgebungen (GRISHMAN/HIRSCHMAN 1978) realisiert werden kann. Vergleichbare Mechanismen prägen auch das linguistische Dokument-Retrieval durch die strukturelle Analyse der natürlichsprachlichen Suchfrage und die sich daran anschließende Referenzform des Text-Retrieval (HALLER/WIELAND 1978).

Linguistische Volltextrepräsentation und -transformation reduziert sich unter einem vereinheitlichenden Blickwinkel auf das Erkennen/Erzeugen von durchaus komplexen syntaktisch definierten Strukturen sowohl mit Blick auf das Referenz- als auch das Fakten-Retrieval. Dazu werden neben Ad-hoc-Grammatiken (etwa in den Systemen DETECT (ROSTEK 1979) oder CONDOR (HALLER/WIELAND 1978)) und deskriptiven Grammatiken (wie in JUDO (FELZMANN et al. 1980)) für das Volltext-Indexing dann im Kontext entwickelterer Formen des linguistischen Information Retrieval - und die automatische Synthese von Datenbanken am Beispiel des 'Linguistic String Project' (SAGER 1978, HIRSCHMAN/GRISHMAN 1977, GRISHMAN/HIRSCHMAN 1978) mag hierfür exemplarisch stehen - strukturalistische Grammatik-Konzeptionen, wie die von HARRIS entwickelte 'String Grammar' (SAGER 1981), eingesetzt. Deren systematisches Semantik-Defizit wird gemildert durch semantisch interpretierte Distributionsklassen, die in eingeschränkten Fachwelten noch recht unproblematisch zu erstellen und zu verwalten und sogar mit dem Konzept der Fachsprachen modellierenden 'sublanguage grammars' (SAGER 1982) theoretisch fundiert sind. Vergleichbare semantisch interpretierbare Konstrukte stehen bei oberflächenorientierten Kasusgrammatiken in Form von Kasusrollen u.ä. zu Verfügung (YOUNG 1973, ROTHKEGEL 1982).

Weit unterhalb dieses analytischen Niveaus rangieren linguistische Beiträge zum automatischen Abstracting (Überblicksdarstellung für diesen Ansatz: PFEIFFER-JÄGER 1980). Unter Umkehrung der vom statistischen Ansatz bekannten Einschlußindikatoren für Sätze eines Dokuments in ein Extrakt (EDMUNDSON 1969) werden beim am weitesten entwickelten linguistischen Abstracting-System ADAM Ausschlußindikatoren für Sätze in einer Negativliste (Umfang: ca. 800 Terms) durch voreingestellte syntaktische (Wortartinformationen) und semantische (Relevanz-, Kontextinformationen usw.) Einschätzungsparameter qualifiziert. Das gemeinsame Auftreten von Listenelementen in einem Satz wird kumulativ (entsprechend der Voreinstellungen) bewertet, so daß entschieden werden kann, ob der

Satz Teil des Extrakts ist oder nicht (RUSH et al. 1971). Die so erzeugten Extrakte werden unter Berücksichtigung kruder strukturalistischer Regularitäten (im Sinne FRIES') zur Verbesserung der Textkohäsion modifiziert (Koordinationstransformationen u.ä.) (MATHIS et al. 1973).

Theoretisch entwickelter sind Abstracting-Konzeptionen, die ihr methodisches Repertoire aus der sowjetischen strukturalistischen Linguistik ableiten. Die (oft nur wenig experimentell abgestützten) Arbeiten setzen den Schwerpunkt der Textanalyse auf Formen der lexikalischen und semantischen Rekurrenz (Anaphora, semantische Kontiguität) sowie der Extension von Konnektoren. Diese Strukturierungsmittel für die Textkohäsion dienen zur Segmentierung kohäsiver Teiltexthe im Sinne einer Textpartitionierung, die als Aufsetzpunkt für Textzusammenfassungen auf unterschiedlichen Kondensierungsniveaus dient (BERZON 1971, BONDARENKO 1976, PRIKHODKO/ SKOROKHODKO 1982). Dieser mehr spekulativ orientierten Forschungsstradition zum automatischen Abstracting kann auch die Erweiterung der 'Funktionalen Satzperspektive' auf eine 'Funktionale Textperspektive' zugeordnet werden (JANOS 1979), deren deskriptive Textstrukturformeln Vorläufer späterer fachtextspezifischer 'story grammars' sind (SALLIS 1978b).

Angesichts der auffälligen Stagnation linguistischer Konzepte für ein automatisches Abstracting erscheint die Hypothese plausibel (KUHLEN 1983), daß die primär oberflächen-syntaktisch geprägten Verfahren der Linguistik Lösungen in semantisch aufwendigen Bereichen der Textanalyse, wie dem Abstracting, nur bis zu einer wenig akzeptablen Leistungsschwelle gestatten (über Erfahrungen in der Anwendungsumgebung von 'Chemical Abstracts Service' mit dem ADAM-System berichten POLLOCK/ZAMORA 1975). Deren Überwindung - das zeigt der Erfolg von unter realen Rahmenbedingungen einsetzbaren Frage-Antwort-Systemen - verlangt offenbar die Einbeziehung starker allgemeiner semantischer und pragmatischer Repräsentationskonstrukte, die auch die theoretische Linguistik bislang nicht anzubieten vermochte.

4.1.4 Der wissensbasierte Ansatz

Der methodische Kern des wissensbasierten Ansatzes zur Volltextverarbeitung besteht aus den Wissensrepräsentations- und -transformationskonzepten der Künstlichen Intelligenz (exemplarisch: HABEL 1982, SCHANK 1982). Die Modellierung etwa von konzeptuellem, prozeduralem oder heuristischem Wissen durch unterschiedlichste Repräsentationsformate (Produktionsregelsysteme, logische Sprachen, semantische Netze, Frames oder Scripts) setzt den Schwerpunkt der wissensbasierten Volltextverarbeitung auf allgemeine Schemata für die Textrepräsentation. Unterschiedliche Typen von Texttransformationen (hier also die Kondensierung) leiten sich in diesem Paradigma unmittelbar aus Operationen über diesen Repräsentationsstrukturen ab, die funktional unabhängig von einzelnen Transformationsvarianten sind. Diese Sicht, Texttrans Transformationsprobleme über allgemeine Textrepräsentationskonstrukte zu lösen, kontrastiert mit dem informatischen und statistischen Ansatz, die beide Transformationsleistungen direkt anstreben mit unmittelbar an die vorgegebene Transformationsform angepaßten inhaltlichen Repräsentationsstrukturen. Diese Einschätzung gilt mittelbar auch für grammatische Konzepte der Linguistik, die zwar allgemeine syntaktische Repräsentationsstrukturen bereitstellen, aber Einschränkungen im Bereich der Transformationsoptionen wegen des semantischen Modellierungsdefizits besitzen. Die Bereitstellung allgemeiner Repräsentationssprachen ermöglicht eine breite Funktionsvielfalt für wissensbasierte Volltext-Informationssysteme dergestalt, daß ein System nicht nur etwa für den Aufbau von Wissensbanken eingesetzt,

sondern mit relativ geringem Aufwand (im Bereich der Transformationsfunktionen) auch als Frage-Antwort-System oder Textzusammenfassungssystem konfiguriert werden kann (SCHANK/RIESBECK 1981, SIMMONS 1984). Insofern sind viele der folgenden Zuordnungen nicht zwingend, charakterisieren aber die Entwicklungsschwerpunkte der einzelnen Systeme.

Wie beim linguistischen Ansatz ist ein Schwerpunkt von Textverstehenssystemen die automatische Synthese von Daten-(hier besser: Wissens)banken aus der Analyse von Volltexten (MONTGOMERY 1983, NISHIDA et al. 1983, REEKER et al. 1983) bzw. weniger ambitiös deren interaktiver Aufbau (HANAKATA 1980). Komplexe semantische Datenstrukturen (Frames, Scripts, Templates usw.) zur Vorstrukturierung des fachspezifischen Weltwissens ersetzen dabei die nach distributionellen Kriterien entwickelten Wortklassenkategorien struktureller Grammatiken (SAGER 1981).

In Form natürlichsprachlicher Anfragen kann dann auf automatisch oder manuell-intellektuell generierte Wissensbestände aus Texten zugegriffen werden: Nach der Übersetzung der Suchfrage in die Form der dem jeweiligen System zugrunde liegenden Wissensrepräsentationssprache und dem Abgleich dieser Struktur mit der Wissenbank kann die Systemantwort im Fakt-Retrieval-Modus (wie in dem aus FRUMP und CYRUS zusammengekoppelten CyFr-System: SCHANK et al. 1981) oder als Spezialfall des Text-Retrieval in Form von einschlägigen Textpassagen präsentiert werden (beim POLYTEXT-System: LOEF 1980, KARLGREN/WALKER 1983; HOBBS et al. 1982).

Als Alternative oder - in abgeschwächter Form - als Ergänzung zum natürlichsprachlichen Mensch-Maschine-Dialog werden graphisch-interaktive Kommunikationsformen eingesetzt, die einen höheren Grad an "Benutzerfreundlichkeit" im Umgang mit den z.T. hoch strukturierten Informationen gewährleisten sollen (vgl. z.B. ZDYBEL et al. 1981).

Diese eher an der Systemoberfläche orientierten Designvarianten für Interfaces werden ergänzt durch Ansätze, die auch den Inhalt des Mensch-Maschine-Dialogs in die Determinierung der Reaktionen des Systems einbeziehen. Im Rahmen von wissensbasierten Retrieval-Interfaces bzw. von Retrieval-Expertensystemen (vgl. z.B. TOU 1982) wird der Retrievalprozeß teilweise modelliert, um den Dialog durch heuristisches Retrievalwissen in Form von Regeln an die situationsspezifischen Erfordernisse zu adaptieren.

Die Arbeiten zum Textinformationsmanagementsystem TOPOGRAPHIC sind diesem Forschungsbereich zuzuordnen.

Das Retrieval von Fakten- bzw. Volltext(passagen) kann man als die Polstellen eines informationellen Kontinuums auffassen, in dem wesentlich flexiblere Formen des Umgangs mit Textinformationen realisiert werden können. Ein Typ ist dabei die Zusammenfassung von Texten im Rahmen von Textkondensierungs- bzw. Abstracting-Systemen. Die Textkondensierungsstrategien sind nahezu ausschließlich auf die Textrepräsentation bezogen, etwa durch Berücksichtigung der Zusammenhängigkeit von Konzeptgruppen in semantischen Netzen (TAYLOR 1974), durch a-priori-Ausdünnung von Repräsentationsstrukturen (Ersetzung vollständiger Scripts durch "sketchy" Scripts in der FRUMP-Komponente von CyFr: SCHANK et al. 1981) oder regelbasierte Parametrisierung der Textkondensierung in Form von benutzerseitig zu spezifizierenden Text- und Zusammenfassungsschemata (FUM et al. 1982). Unter der hier zugrunde liegenden streng repräsentationellen Sicht könnte man auch die Generierung von Texten aus formatierten Datenstrukturen einer (konventionellen numerischen) Datenbank (KUKICH 1983) als einen Kondensierungsprozeß betrachten, in dem numerische Manipulationen und Relevanzregeln der dabei festgestellten Trends die

Informationen, die zur Textausgabe kommen, steuern. Aus systematischen Gründen aber wird der Aspekt der Textgenerierung hier nicht weiter verfolgt (vgl. den Überblick in MANN et al. 1981).

Zunehmend in den Vordergrund treten auch im Bereich wissensbasierter Volltext-Informationssysteme Aspekte der (kognitiv) adäquaten Speicherung und des Zugriffs auf Wissenssegmente, etwa durch konzeptuell motivierte Indexstrukturen (KOLODNER 1984; vgl. auch die CYRUS-Komponente von CyFr: SCHANK et al. 1981). Dabei werden Wissensbanken nicht als passive Speichermedien verstanden, sondern kontrollieren in Interaktion mit den Textanalyse-Mechanismen die Einziehung neuer Konzepte bei der Wissensakquisition und steuern die Tiefe des Textverstehens, etwa durch Bewertung des Neuigkeitswerts von Informationen in einem analysierten Text (LEBOWITZ 1983).

Im Schnittfeld dieser Systementwicklungen ist die Arbeit am Textkondensierungssystem TOPIC anzusiedeln.

4.1.5 Der kognitive Ansatz

Während Ansätze der kognitiven Psycholinguistik, das menschliche Verstehen von Texten als Modell einer Volltextverarbeitung (vgl. exemplarisch KINTSCH/VAN DIJK 1978) in Informationssystemen zu betrachten, bisher kaum in operationalisierten Verfahren resultieren, finden Ergebnisse von Untersuchungen, die den Menschen als "Komponente" eines Gesamtsystems "Mensch/Informationssystem" betrachten, zunehmend Eingang in das Design (zumindest der Schnittstellen) von Informationssystemen. Im folgenden werden verschiedene Aspekte kognitiver "human factors"-Forschung in bezug auf ihre Relevanz für Volltextverarbeitungssysteme dargestellt.

Einen umfassenden Überblick zur Situation der "human factors"-Forschung im Bereich des Mensch-Maschine-Dialogs vermitteln RAMSEY/GRIMES 1983, die weitgehend einer von MORAN 1981 vorgeschlagenen Systematisierung der verschiedenen Ansätze folgen.

Relevant für das Oberflächendesign von Informationssystemen, die Information (z.T.) in textueller Form vermitteln, sind empirische Resultate der Perzeptionsforschung (z.B. KENNEDY 1983, WAERN/ROLLENHAGEN 1983, BEVAN 1981). Dabei werden Fragen zum ergonomischen Layout der Schnittstelle, wie Positionierung und Präsentationsgeschwindigkeit von Textfragmenten auf dem Bildschirm anhand empirischer Tests angegangen und z.T. gelöst.

Auf der Ebene der menschlichen Informationsverarbeitung werden im Rahmen der Retrievalforschung Suchstrategien (vgl. ROUSE/ROUSE 1984) und -taktiken (BATES 1979) analysiert. Resultate dieser Forschungen bilden die Basis für die explizite Modellierung des Retrievalverhaltens in Form heuristischer Regeln als Teil eines wissensbasierten Interface für Volltextinformationssysteme. In die Konzeption einer solchen Schnittstelle müssen schließlich auch die Erfahrungen mit bestehenden Retrievalsystemen einfließen, deren Features auf ihre kognitive Adäquanz zu prüfen sind. Ansätze hierzu finden sich z.B. in INGWERSEN 1984.

Bislang sind keine Systeme implementiert, die Erkenntnisse kognitiver Untersuchungen in die Systemstruktur (der Benutzerschnittstelle) eines Volltextinformationssystems integrieren. Im Bereich der Expertensysteme vollzieht sich zur Zeit jedoch eine Entwicklung, die als Analogon zu einer Einbeziehung von Retrievalwissen in die Dialogführung anzusehen ist. Während z.B. das System ONCOCIN (LANGLOTZ/SHORTLIFFE 1983) Therapiepläne von Medizinern einer auf heuristischen Regeln beruhenden "Kritik"

unterwirft, modelliert das System SHRINK (KOLODNER 1983) psychiatrische Diagnose- und Therapieplanung.

Die Formulierung und Implementierung von Retrievalheuristiken im Teilprojekt TOPOGRAPHIC findet im Kontext dieser Ansätze statt.

L i t e r a t u r v e r z e i c h n i s

- ARTANDI, S. (1969)
Computer Indexing of Medical Articles: Project MEDICO. In: Journal of Documentation 25. 1969, No.3, pp.214-223.
- ATTAR, R. / A.S. FRAENKEL (1977)
Local Feedback in Full-Text Retrieval Systems. In: Journal of the ACM 24. 1977, No.3, pp.397-417.
- BATES, M. (1979)
Information Search Tactics. In: Journal of the ASIS 30. 1979, pp.205-214
- BERNSTEIN, L.M. / R.E. WILLIAMSON (1984)
Testing of a Natural Language Retrieval System for a Full Text Knowledge Base. In: Journal of the ASIS 35. 1984, No.4, pp.235-247.
- BERZON, V.E. (1971)
An Approach to Problems of Automatic Abstracting and Automatic Text Contraction. In: Automatic Documentation and Mathematical Linguistics 5. 1971, No.4, pp.11-20.
- BEVAN, N. (1981)
Is There an Optimum Speed for Presenting Text on a VDU? In: Int. J. Man-Machine-Studies 14. 1981, pp.59-76
- BONDARENKO, G.V. (1976)
The Use of Formal Text Structure Analysis Methods for Automatic Indexing and Abstracting. In: Automatic Documentation and Mathematical Linguistics 10. 1976, No.2, pp.62-73.
- BOREHAM, J. / B. NIBLETT (1976)
Classification of Legal Texts by Computer. In: Information Processing & Management 12. 1976, No.2, pp.125-132.
- CROFT, W.B. / J. ROCHFORD (1981)
Evaluation of Text Indexing and Retrieval in the Office Environment. Amherst/MA: Computer and Information Science, University of Massachusetts at Amherst, Dec 1981 (= COINS Technical Report 81-27).
- DILLON, M. / L.K. McDONALD (1983)
Fully Automatic Book Indexing. In: Journal of Documentation 39. 1983, No.3, pp.135-154.
- EDMUNDSON, H.P. (1969)
New Methods in Automatic Extracting. In: Journal of the ACM 16. 1969, No.2, pp.264-285.

- FELZMANN, F.W. et al. (1980)
JUDO - Juristische Dokumentanalyse (Berichtszeitraum: 1.7.77 - 31.12.79). 2 Vols. Regensburg: Univ. Regensburg, Abt.für Nichtnumerische Datenverarbeitung, Mai 1980.
- FREI, H.P. / JAUSLIN, J.F. (1983)
Graphical Presentation of Information and Services: A User-oriented Interface. In: Information Technology: Research and Development 2. 1983, pp.23-42.
- FUM, D. / G. GUIDA / C. TASSO (1982)
Forward and Backward Reasoning in Automatic Abstracting. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia, 1982, pp.83-88.
- GRISHMAN, R. / L. HIRSCHMAN (1978)
Question Answering from Natural Language Medical Data Bases. In: Artificial Intelligence 11. 1978, pp.25-43.
- HABEL, C. (1982)
Textverstehende Systeme: ein Beispiel für die Beziehung Informatik - Kognitionswissenschaft. In: J. Nehmer (ed): Proceedings der 12. Jahrestagung der GI. Kaiserslautern, 5.-7. Okt. 1982. Berlin etc.: Springer, 1982, pp.372-391.
- HAHN, U. (1983)
Leistungsfähigkeit und Grenzen externer publizistischer Datenbanken und deren Verbesserung durch informationslinguistische Software. In: Medienarchive in Gegenwart und Zukunft. Protokoll der 39. und 40. Tagung der Fachgruppe Presse-, Rundfunk- und Filmarchivare im Verein deutscher Archivare. Heilbronn, Oktober 1981 und Köln, Mai 1982. München etc.: Saur, 1983, pp.109-135 (= Mediendokumentation 4).
- HALLER, J. / U. WIELAND (1978)
Die Erschließung natürlichsprachlicher Information im Informationssystem CONDOR. In: Nachrichten für Dokumentation 29. 1978, Nos.4-5, pp.177-183.
- HANAKATA, K. (1980)
An Intelligent Digester for Interactive Text Processings. In: COLING 80: Proceedings of the 8th International Conference on Computational Linguistics. Sept.30-Oct.4, 1980, Tokyo. Tokyo 1980, pp.504-511.
- HIRSCHMAN, L. / R. GRISHMAN (1977)
Fact Retrieval from Natural Language Medical Records. In: D.B. Shires / H. Wolf (eds): MEDINFO 77: Proceedings of the 2nd Conference on Medical Informatics. Toronto, August 8-12, 1977. Amsterdam etc.: North-Holland, 1977, pp.247-251.
- HOBBS, J.R. / D.E. WALKER / R.A. AMSLER (1982)
Natural Language Access to Structured Text. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia, 1982, pp.127-132.
- HOLLAAR, L.A. (1983)
Hardware Systems for Text Information Retrieval. In: Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. National Library of Medicine,

Bethesda, Maryland, June 6-8, 1983 (= SIGIR Forum 17. 1983, No.4), pp.3-9.

INGWERSEN, P. (1984)

Online Man-Machine Interaction Facilities: A Cognitive View. In: H.J. Dietschmann (ed): Representation and Exchange of Knowledge as a Basis of Information Processes, Proceedings IRFIS 5. Heidelberg, September 5-7, 1983. Amsterdam etc.: North-Holland, 1984, pp.325-358.

JANOS, J. (1979)

Theory of Functional Sentence Perspective and its Application for the Purposes of Automatic Extracting. In: Information Processing and Management 15. 1979, No.1, pp.19-25.

KARASEV, S.A. (1974)

A Study of Sentence Significance Criteria in Automatic Quasi-abstracting. In: Automatic Documentation and Mathematical Linguistics 8. 1974, No.4, pp.79-87.

KARLGRÉN, H. / D.E. WALKER (1983)

The Polytext System: a New Design for a Text Retrieval System. In: F. Kiefer (ed): Questions and Answers. Dordrecht etc.: Reidel, 1983, pp.273-294 (= Linguistic Calculation 1).

KENNEDY, A. (1983)

On Looking into Space. In: K. Rayner (ed): Eye Movements in Reading. Perceptual and Language Processing. New York etc.: 1983, pp.237-250.

KINTSCH, W./ T.A. VAN DIJK (1978)

Toward a Model of Text Comprehension and Production. In: Psychological Review 85. 1978, No.5, pp.363-394.

KOLODNER, J.L. (1983)

Towards an Understanding of the Role of Experience in the Evolution from Novice to Expert. In: Int. J. Man-Machine-Studies 19. 1983, pp.497-518.

KOLODNER, J.L. (1984)

Retrieval and Organizational Strategies in Conceptual Memory: a Computer Model. Hillsdale/NJ etc.: Erlbaum, 1984.

KUHLEN, R. (1983)

Volltextanalyse zum Zwecke des Abstracting. In: Deutscher Dokumentar-tag 1982: Fachinformation im Zeitalter der Informationsindustrie. Lübeck-Travemünde, 27.9.-30.9.1982. München etc.: Saur, 1983, pp.300-312.

KUKICH, K. (1983)

Knowledge-Based Report Generation: A Technique for Automatically Generating Natural Language Reports from Databases. In: Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. National Library of Medicine, Bethesda, Maryland, June 6-8, 1983, (= SIGIR Forum 17. 1983, No.4), pp.246-250.

LANGLOTZ, C.P. / E.H. SHORTLIFFE (1983)

Adapting a Consultation System to Critique User Plans. In: Int. J. Man-Machine-Studies 19. 1983, pp.479-496.

- LEBOWITZ, M. (1983)
Intelligent Information Systems. In: Proceedings of the 6th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. National Library of Medicine, Bethesda, Maryland, June 6-8, 1983, (= SIGIR Forum 17. 1983, No.4), pp.25-30.
- LOEF, S. (1980)
The POLYTEXT/ARBIT Demonstration System. Umea/Sweden: Foersvarets Forskningsanstalt, Sept 1980. (= FOA 4 rapport, C 40121-M7)
- MACLEOD, I.A. (1981)
A Data Base Management System for Document Retrieval Applications. In: Information Systems 6. 1981, No.2, pp.131-137.
- MAH, C.P. / R.J. D'AMORE (1983)
Complete Statistical Indexing of Text by Overlapping Word Fragments. In: SIGIR Forum 17. 1983, No.3, pp.6-16.
- MANN, W.C. / M. BATES / B.J. GROSZ / D.D. McDONALD / K.R. McKEOWN / W.R. SWARTOUT (1981)
Text Generation: The State of the Art and Literature. Marina del Ray/CA: Univ. of Southern California, Information Sciences Institute, Dec. 1981. (= ISI/RR-81-101)
- MARON, M.E. (1982)
Problems with Full-Text Searching. In: Office Automation Conference 1982 Digest. Moscone Center, San Francisco, April 5-7, 1982. o.O.: AFIPS, 1982, pp.379-384.
- MARTIN, J.M. et al. (1977)
Automatic Printing in Clear Language of Complete Medical Observations. In: D.B. Shires / H. Wolf (eds): MEDINFO 77: Proceedings of the 2nd World Conference on Medical Informatics. Toronto, August 8-12, 1977. Amsterdam etc.: North-Holland, 1977, pp.293-295.
- MATHIS, B.A. / J.E. RUSH / C.E. YOUNG (1973)
Improvement of Automatic Abstracts by the Use of Structural Analysis. In: Journal of the ASIS 24. 1973, No.2, pp.101-109.
- MORAN, T.P. (1981)
An Applied Psychology of the User. In: ACM Computing Surveys 13. 1981, No.1, pp.2-11.
- MONTGOMERY, C.A. (1983)
Distinguishing Fact from Opinion and Events from Meta-Events. In: Proceedings of the Conference on Applied Natural Language Processing. 1-3 February 1983, Santa Monica, CA. o.O.: ACL, 1983, pp.55-61.
- NISHIDA, T. / A. KOSAKA / S. DOSHITA (1983)
Towards Knowledge Acquisition from Natural Language Documents: Automatic Model Construction from Hardware Manuals. In: IJCAI 83: Proceedings of the 8th International Joint Conference on Artificial Intelligence. 8-12 August 1983, Karlsruhe, West Germany. Vol.1. Los Altos/CA: Kaufmann, 1983, pp.482-486.
- O'CONNOR, J. (1973)
Text Searching Retrieval of Answer-Sentences and other Answer-Passages. In: Journal of the ASIS 24. 1973, No.6, pp.445-460.

- PFEIFFER-JÄGER, G. (1980)
Referat und Referieren: Linguistische Beiträge zu ihrer Applikation in der Information und Dokumentation. In: Germanistische Linguistik 1980 (1982), Nos.1-2, pp.1-180.
- POLLOCK, J.J. / A. ZAMORA (1975)
Automatic Abstracting Research at Chemical Abstracts Service. In: Journal of Chemical Information and Computer Sciences 15. 1975, No.4, pp.226-233.
- PRIKHODKO, S.M. / E.F. SKOROKHODKO (1982)
Automatic Abstracting from Analysis of Links between Phrases. In: Automatic Documentation and Mathematical Linguistics 16. 1982, No.1, pp.55-65.
- RABITTI, F. / J. ZIZKA (1984)
Evaluation of Access Methods to Text Documents in Office Systems. In: C.J. van Rijsbergen (ed): Research and Development in Information Retrieval. Proceedings of the 3rd Joint BCS and ACM Symposium, King's College, Cambridge, 2-6 July 1984. Cambridge etc.: Cambridge Univ. Pr., 1984, pp.21-40.
- RAMSEY, H.R. / J.D. GRIMES (1983)
Human Factors in Interactive Computer Dialog. In: Annual Review of Information Science and Technology, 18. 1983, pp.29-59.
- REEKER, L.H. / E.M. ZAMORA / P.E. BLOWER (1983)
Specialized Information Extraction: Automatic Chemical Reaction Coding from English Descriptions. In: Proceedings of the Conference on Applied Natural Language Processing. 1-3 February 1983, Santa Monica, CA. o.o.: ACL, 1983, pp.109-116.
- RIJSBERGEN, C.J. van (1979)
Information Retrieval. London etc.: Butterworths, 1979, 2nd ed.
- 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, 1979, pp.251-282.
- ROTHKEGEL, A. (1982)
Zur Analyse von Textinformationen. In: H. Fix et al. (eds): Sprachen und Computer. Dudweiler: AQ, 1982, pp.365-383 (= Sprachwissenschaft - Computerlinguistik 9).
- ROUSE, W.B. / S.H. ROUSE (1984)
Human Information Seeking and Design of Information Systems. In: Information Processing & Management 20. 1984, No.1, pp.129-138.
- RUSH, J.E. / R. SALVADOR / A. ZAMORA (1971)
Automatic Abstracting and Indexing. Pt.2: Production of Indicative Abstracts by Application of Contextual Inference and Syntactic Coherence Criteria. In: Journal of the ASIS 22. 1971, No.4, pp.260-274.
- SAGER, N. (1978)
Natural Language Information Formatting: The Automatic Conversion of Texts to a Structured Data Base. In: M.C. Yovits (ed): Advances in Computers. Vol. 17. New York etc: Academic Pr., 1978, pp.89-162.

- SAGER, N. (1981)
Natural Language Information Processing: A Computer Grammar of English and Its Applications. Reading/MA: Addison-Wesley, 1981.
- SAGER, N. (1982)
Syntactic Formatting of Science Information. In: R. Kittredge / J. Lehrberger (eds): Sublanguage: Studies of Language in Restricted Semantic Domains. Berlin etc.: de Gruyter, 1982, pp.9-26.
- SALLIS, P.J. (1978a)
Text Processing: a Matter of Definition or Application? In: Program 12. 1978, No.4, pp.185-187.
- SALLIS, P.J. (1978b)
A Partial-Parsing Algorithm for Natural Language Text Using a Simple Grammar for Arguments. In: ALLC Bulletin 6. 1978, pp.170-176.
- SCHANK, R.C. (1982)
Representing Meaning: an Artificial Intelligence Perspective. In: S. Allen (ed): Text Processing: Text Analysis and Generation. Text Typology and Attribution. Proceedings of the 'Nobel Symposium 51'. Stockholm: Almqvist & Wiksell, 1982, pp.25-63 (= Data Linguistica 16).
- SCHANK, R.C. / J.L. KOLODNER / G. DeJONG (1981)
Conceptual Information Retrieval. In: R.N. Oddy / S.E. Robertson / C.J. van Rijsbergen / P.W. Williams (eds): Information Retrieval Research. London etc.: Butterworths, 1981, pp.94-116.
- SCHANK, R.C. / C.K. RIESBECK (eds) (1981)
Inside Computer Understanding: Five Programs plus Miniatures. Hillsdale/NJ: Erlbaum, 1981.
- SCHEK, H.-J. (1981)
Methods for the Administration of Textual Data in Database Systems. In: R.N. Oddy et al. (eds): Information Retrieval Research. London: Butterworths, 1981, pp.218-235.
- SIMMONS, R.F. (1984)
Computations from the English: a Procedural Logic Approach for Representing and Understanding English Texts. Englewood Cliffs/NJ: Prentice-Hall, 1984.
- TAYLOR, S.L. (1974)
Automatic Abstracting by Applying Graphical Techniques to Semantic Networks. Evanston/Illinois: Northwestern University, Computer Science, (Aug) 1974, (Ph.D.Thesis).
- TOLIVER, D.E. (1982)
OL'SAM: An Intelligent Front-End for Bibliographic Information Retrieval. In: Information Technology and Libraries 1. 1982, No.4, 1982, pp.317-326.
- TOU, F.N. et al. (1982)
RABBIT: An Intelligent Database Assistant. In: AAAI-82: Proceedings of the Second National Conference on Artificial Intelligence. Pittsburgh, Pennsylvania, August 1982. Los Altos, CA: Kaufmann, 1982, pp.314-318.

- WAERN, Y. / C. ROLLENHAGEN (1983)
Reading Texts from Visual Display Units (VDUs). In: Int. J. Man-Machine-Studies 18. 1983, pp.441-465.
- WYLLYS, R.E. (1967)
Extracting and Abstracting by Computer. In: H. Borko (ed): Automated Language Processing. New York etc.: Wiley, 1967, pp.127-179.
- YOUNG, C.E. (1973)
Development of Language Analysis Procedures with Application to Automatic Indexing. o.O.: Ohio State University, Computer and Information Science Research Center, 1973. (Ph.D.Thesis)
- ZDYBEL, F. / N. GREENFELD / M. YONKE / J. GIBBONS (1981)
An Information Presentation System. In: Proc. of the 7th Int. Joint Conf. on AI. Vancouver 1981, pp.978-984.

4.2 Ein integrales Konzept der Textanalyse und des Textinformationsmanagement in einem wissensbasierten Volltext-Informationssystem

4.2.1 Textanalyse in TOPIC

Der Analyse und Repräsentation von Volltexten im Textkondensierungssystem TOPIC liegen auf Texte besonders abgestimmte Analyseverfahren zugrunde, und zwar - wie aus dem detaillierten Arbeitsplan (Kap. 3) ersichtlich -

- (1) zur Überführung von Texten in eine Textrepräsentationsstruktur (Text-Parsing)
- (2) zur adäquaten Repräsentation und Verwaltung des in den Texten enthaltenen Wissens (Textrepräsentation)
- (3) Transformationsmethoden zur Umformung von Textwissens- in Kondensatstrukturen (Textkondensierung)

4.2.1.1 Text-Parsing

Mit der Analyse von Volltexten ist linguistisch der Ebenenwechsel verbunden vom Einzelsatz, der nach wie vor das bevorzugte Beschreibungsobjekt auch neuerer Grammatik-Formalismen (wie LFG, GPSG, DCG) ist, auf den Text. Dies hat z.T. erhebliche Konsequenzen in bezug auf die zu beschreibenden sprachlichen Phänomene (vgl. z.B. PETIOEFI 1979) und darüber hinaus auf die Organisation des Beschreibungsmechanismus, wie die Entwicklung unterschiedlicher Entwürfe zu Text- (statt Satz-) Grammatiken zeigt (einen groben Überblick über linguistische, logische und mathematische Formalismen für Textgrammatiken gibt DIJK 1973). Sind textspezifische Phänomene ausschließlich nur durch Textgrammatiken zu beschreiben, so folgt für die Anwendung natürlichsprachlicher Systeme auf Texte, daß deren Parser prinzipiell nicht mehr satzorientiert sein können.

Neben diesem analytischen Argument haben die Entwicklung des Text-Parsing-Mechanismus in TOPIC auch mehr pragmatische Gesichtspunkte geprägt. Zum einen die Überlegung, daß - ausgehend vom informationellen Analyseziel "indikatives Textkondensat" - nicht die Notwendigkeit einer umfassenden, auf Vollständigkeit ausgerichteten Textanalyse besteht. Dies begründet sich zum einen aus Erfahrungen mit Information-Retrieval-Systemen, deren lexikalisch orientierte Analyseverfahren die in TOPIC

entwickelten Instrumente im Sinne eines partiellen Text-Parsing konzeptionell beeinflusst haben. Die dabei entwickelte Form einer konzept- (statt syntax-) orientierten Analyse unterstützt weiter die Robustheit und Flexibilität des Parsers, die für die Verarbeitung "realer" Texte (im Sinne der "realistic language comprehension" (RIESBECK 1982)) unverzichtbar ist. Für dieses Vorgehen läßt sich andererseits aber auch kognitive Plausibilität reklamieren, denn das menschliche Textprozessierungsverhalten, insbesondere bei dem Informationsproblem, sich eine hinreichende Übersicht über die Inhaltsschwerpunkte von Texten zu verschaffen, ähnelt wenig einer gleichmäßig "tiefen", vollständigen Verarbeitung jedes Textelements, sondern mehr einem "überfliegen", "Querlesen", m.a.W. einem extrem oberflächlichen Textverstehen, das aber hinreicht, um den groben Inhaltskern des Textes zu erschließen (THIBADEAU et al. 1982). In TOPIC wird dieses Argument nicht verwandt, um für die angestrebten Implementierungslösungen zugleich kognitive Plausibilität/ Adäquatheit zu reklamieren, sondern es wird als Hinweis auf die Entwicklungsbedürftigkeit text- und funktionsspezifischer Analysemechanismen interpretiert.

Aus diesen Überlegungen sind zwei Grundanforderungen an den Text-Parser des TOPIC-Systems abgeleitet worden:

- die Modellierung von Texteigenschaften beim Parsing
- die heuristische Voreinstellung des Parsers auf eine oberflächliche Textanalyse

Erfahrungen mit Systemen, die "flaches" Textverstehen simulieren, zeigen, daß dies prinzipiell über die ausgeprägte Integration von lexikalisch bzw. phrasen-orientierter und konzeptueller Analyse der Texte (DeJONG 1979, SHANK et al. 1980, CATER 1980, COWIE 1983) operationalisiert wird. In Form eines Wortexperten-Systems (SMALL 1980) steht ein dieser "Wort"-Orientierung exakt entsprechendes Parsing-Instrument in TOPIC zur Verfügung. Partielles Text-Parsing resultiert dann aus der gezielten Begrenzung der für die Analyse verfügbaren Kollektion von Wortexperten (und Wissensrepräsentationsstrukturen).

Die Modellierung von Textkohäsions- und -kohärenzeigenschaften, dem eigentlichen Schwerpunkt der Arbeiten zum Text-Parsing im zweiten Projektabschnitt, unterstützt die Organisation von und den Zugriff auf Textrepräsentationsstrukturen in den Welt- und Textwissensbanken des TOPIC-Systems und steuert wesentlich die Textkondensierung (bzw. in der Umkehrung das Text-Retrieval durch TOPOGRAPHIC). Dies scheint jedoch nicht allein für die begrenzten Ziele eines Textkondensierungssystems unverzichtbar, sondern für jede Form von Textanalysestextsystem gültig zu sein. Mit Textstrukturen werden nämlich auch Organisationsinformationen für Textwissen vermittelt, deren repräsentationstechnische Nichterkennung/-berücksichtigung zu unterstrukturierten Wissensbasen führt.

Sprachliche Mittel der Textkohäsion legen die Expansion eines Textthemas und damit seine Abgrenzung gegenüber anderen Textthemen fest. Dabei werden als grundlegende sprachliche Kohäsionsmittel unterschieden: Referenz und Anaphora, Substitution und Ellipse, Konnektoren, lexikalische Kohäsion/Kollokation (HALLIDAY/HASAN 1976). In neueren textwissenschaftlichen Arbeiten werden dem Bereich der Textkohäsionsphänomene auch Inferenzen zugezählt (CROTHERS 1979).

Vorarbeiten zu Aspekten der Textkohäsion in natürlichsprachlichen Systemen konzentrieren sich auf die Behandlung von Referenzen und Anaphora (einen Überblick über Strategien zur Anaphoraauflösung gibt HIRST 1981) bzw. Inferenzen (z.B. CULLINGFORD 1979). Besonders auf Framesprachen abge-

stellte Modellierungen behandeln Aspekte der Referenz (ROSENBERG/ROBERTS 1979) und der lexikalischen Kohäsion/Kollokation (ROSENBERG 1980). Bei der Behandlung der Konnektive stehen dagegen umfassenden textlinguistischen Vorarbeiten - strukturellen Typologien (z.B. LONGACRE 1983) oder textgrammatischen Modellierungsversuchen (z.B. DIJK 1977) - nur vereinzelt konstruktive Umsetzungen in natürlichsprachlichen Systemen gegenüber. Dies gilt weniger für konjunktive und disjunktive Konnektoren als vielmehr für die seltene Behandlung temporaler, adversativer, kausaler und konzessiver u.ä. Konnektoren (ROESNER/LAUBSCH 1982, ROLLINGER 1984). Dabei sind die heuristischen Effekte dieser expliziten sprachlichen Indikatoren auf den Verstehensprozeß und damit die Organisation der Wissensrepräsentationsstrukturen eines Texts längst erkannt (CLIPPINGER/McDONALD 1983): Sie strukturieren die aus der Satzanalyse ermittelten einzelnen Propositionen im Sinne der Argumentationslinie im Text und verknüpfen sie damit im Sinne der Textkohäsion (WINTER 1978, COHEN 1984).

Die hier unter Textkohärenz subsumierten Arbeiten beziehen sich durchweg auf textstrukturell eingeführte Kohärenz. Textpragmatische Formen der Kohärenz, wie sie in Form von Zielprojektionen und Planverfolgungsstrategien (SCHANK/ABELSON 1977) oder Interessensmodellen (WILENSKY 1982) usf. besonders für das Verstehen narrativer Texte ein zentraler Forschungsaspekt sind, können für die Analyse von Fachtexten (insbesondere für die Zwecke der Textkondensierung) vernachlässigt werden.

Unter den textstrukturellen Kohärenzmustern sind die grundlegenden Muster thematischer Progression, mit denen Gesetzmäßigkeiten der kontinuierlichen Expansion eines Themas bzw. des Themenwechsels in Texten beschrieben werden (DANES 1974, GRIMES 1978), linguistisch am besten untersucht. Eines der wenigen konstruktiven Experimente, thematische Progressionsformen im Rahmen von Textverstehenssystemen zu modellieren, beschreibt CRITZ 1982, der zudem auch einen Frame-Ansatz verfolgt.

Neben diesem Grundtypus sind eine Vielfalt funktionaler Kohärenzrelationen für Texte vorgeschlagen worden: Kontrast- bzw. Kompatibilitätsbeziehungen zwischen benachbarten Textsegmenten, "logische" Beziehungen zwischen den einzelnen Segmenten (Kausalität, Instrumentalität, Konditionalität) sowie an rhetorische Grundmuster angepaßte Relationen, etwa der Übergang vom Generellen zum Spezifischen, vom Hypothetischen zum Realen u.ä. (DEA/BELKIN 1978, HOBBS 1978, 1979, DIJK 1980a, MANN 1984). Realisierungen dieser textlinguistischen Konzepte in natürlichsprachlichen Systemen existieren nur vereinzelt (HOBBS 1982) mit einem Schwerpunkt auf der Behandlung zeitlicher Kohärenz von Textsegmenten (HIRSCHMAN/STORY 1981, GÜNTHER et al. 1983). Der für das Textverstehen relevante Aspekt dieser (binären) Kohärenzrelationen besteht im Aufbau von gezielten Erwartungen in bezug auf die Form des Textverlaufs: Tritt ein Element einer Kohärenzrelation auf, ist im (un)mittelbaren Kontext das Auftreten des komplementären Elements der Kohärenzrelation zu erwarten (DEA/BELKIN 1978). Unterschiedliche Formen von Verschachtelungen dieser Relationen unterstützen so die kohärente Organisation inhaltlich kohäsiver Textsegmente im Sinne der textuellen Wohlgeformtheit eines Texts. Die meisten dieser Regularitäten können auf die Semantik von Konnektoren oder auf Relationeneigenschaften von Konzepten zurückgeführt werden. Dies erklärt, warum in TOPIC der Text-Parser die Hauptlast der Kohärenzverarbeitung trägt, im Unterschied etwa zu mehr wissensrepräsentationstechnischen Lösungen (HOBBS 1977).

Mit der Postulierung von Kohärenzrelationen wird ein relativ flexibles Modell für die Textkohärenz verfolgt, dem das eher statische Konzept von global vorstrukturierten Bildungsmustern für Texte gegenübersteht, wie es

in Form von 'story grammars' (RUMELHART 1975), 'superstructures' (DIJK 1980b) oder Fachtextgrammatiken (SALLIS 1978) formuliert wird. Die bislang verfügbaren Grammatikfragmente erweisen sich zwar als prototypisch einleuchtende Beschreibungsformen, besitzen aber bislang für die Analyse von realen Texten zu wenig Flexibilität. Implementierungen von Story Grammars in Textverstehenssystemen sind daher auch relativ selten (CORREIRA 1980).

L i t e r a t u r v e r z e i c h n i s

CATER, A.W.S. (1980)

Analysing English Text: a Nondeterministic Approach with Limited Memory. In: AISB-80: Conference Proceedings. Amsterdam, 1-4 July, 1980. Society for the Study of Artificial Intelligence and the Simulation of Behaviour, 1980, pp.CATER-1-CATER-14.

CLIPPINGER, J.H. / D.D. McDONALD (1983)

Why Good Writing is Easier to Understand. In: IJCAI-83: Proceedings of the 8th International Joint Conference on Artificial Intelligence. 8-12 August 1983, Karlsruhe, West Germany. Vol.2. Los Altos/CA: Kaufmann, 1983, pp.730-732.

COHEN, R. (1984)

A Computational Theory of the Function of Clue Words in Argument Understanding. In: COLING 84: Proceedings of the 10th International Conference on Computational Linguistics & 22nd Annual Meeting of the Association for Computational Linguistics. 2-6 July 1984, Stanford University, California. ACL 1984, pp.251-258.

CORREIRA, A. (1980)

Computing Story Trees. In: American Journal of Computational Linguistics 6. 1980, Nos.3-4, pp.135-149.

COWIE, J.R. (1983)

Automatic Analysis of Descriptive Texts. In: Proceedings of the Conference on Applied Natural Language Processing. 1-3 February 1983, Santa Monica, CA. o.O.: ACL, 1983, pp.117-123.

CRITZ, J.T. (1982)

Frame Based Recognition of Theme Continuity. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia, 1982, pp.71-75.

CROTHERS, E.J. (1979)

Paragraph Structure Inference. Norwood/NJ: Ablex, 1979.

CULLINGFORD, R. (1979)

Pattern-Matching and Inference in Story Understanding. In: Discourse Processes 2. 1979, No.4, pp.319-334.

DANES, F. (1974)

Functional Sentence Perspective and the Organization of the Text. In: F. Danes (ed): Papers on Functional Sentence Perspective. Prague: Academia, 1974, pp.106-128.

- DEA, W. / N.J. BELKIN (1978)
Beyond the Sentence: Clause Relations and Textual Analysis. In:
Informatics 3: Proceedings of a Conference held by the Aslib
Co-ordinate Indexing Group. 2-4 April 1975, Emmanuel College,
Cambridge. London: Aslib, 1978, pp.67-84.
- DeJONG, G.F. (1979)
Skimming Stories in Real Time: an Experiment in Integrated Under-
standing. Yale University, May 1979. (Ph.D.Thesis)
- DIJK, T.A. van (1973)
Models for Text Grammars. In: Linguistics 1973, No.105, pp.35-68.
- DIJK, T.A. van (1977)
Connectives in Text Grammar and Text Logic. In: T.A. van Dijk / J.S.
Petoefi (eds): Grammars and Descriptions. Berlin etc.: de Gruyter,
1977, pp.11-63 (= Research in Text Theory 1).
- DIJK, T.A. van (1980a)
Text and Context: Explorations in the Semantics and Pragmatics of
Discourse. London etc.: Longman, (1977) 1980, pp.93-129 (= Ch.4).
- DIJK, T.A. van (1980b)
Macrostructures: an Interdisciplinary Study of Global Structures in
Discourse, Interaction, and Cognition. Hillsdale/NJ: Erlbaum, 1980,
pp.107-132 (= Ch. 3).
- GRIMES, J.E. (1978)
Topic Levels. In: TINLAP-2: Theoretical Issues in Natural Language
Processing-2. University of Illinois at Urbana-Champaign, July 25-27,
1978. New York: ACM, 1978, pp.104-108.
- GÜNTHER, S. / C. HABEL / C.-R. ROLLINGER (1983)
Ereignisnetze: Zeitnetze und Referentielle Netze. In: Linguistische
Berichte 1983, No.88 (Dez), pp.37-55.
- HALLIDAY, M.A.K. / R. HASAN (1976)
Cohesion in English. London: Longman, 1976.
- HIRSCHMAN, L. / G. STORY (1981)
Representing Implicit and Explicit Time Relations in Narrative. In:
IJCAI-81: Proceedings of the 7th International Joint Conference on
Artificial Intelligence. 24-28 August 1981, University of British
Columbia, Vancouver/BC, Canada. Vol.1. Menlo Park/CA: AAAI, 1981,
pp.289-295.
- HIRST, G. (1981)
Discourse-Oriented Anaphora Resolution in Natural Language
Understanding: a Review. In: American Journal of Computational
Linguistics 7. 1981, No.2, pp.85-98.
- HOBBS, J.R. (1977)
Coherence and Interpretation in English Texts. In: IJCAI-77:
Proceedings of the 5th International Joint Conference on Artificial
Intelligence. Massachusetts Institute of Technology, Cambridge,
Massachusetts, USA, August 22-25, 1977. Vol.1. 1977, pp.110-116.
- HOBBS, J.R. (1978)
Why is Discourse Coherent? Menlo Park/CA: SRI International, Nov 1978
(= SRI Technical Note 176).

- HOBBS, J.R. (1979)
Coherence and Coreference. In: Cognitive Science 3. 1979, No.1,
pp.67-90.
- HOBBS, J.R. (1982)
Towards an Understanding of Coherence in Discourse. In: W.G. Lehnert /
M.H. Ringle (eds): Strategies for Natural Language Processing.
Hillsdale/NJ etc.: Erlbaum, 1982, pp.223-243.
- LONGACRE, R.E. (1983)
The Grammar of Discourse. New York etc.: Plenum Pr., 1983, pp.77-149
(= Ch.3).
- MANN, W.C. (1984)
Discourse Structures for Text Generation. In: COLING 84: Proceedings
of the 10th International Conference on Computational Linguistics &
22nd Annual Meeting of the Association for Computational Linguistics.
2-6 July 1984, Stanford University, California. ACL 1984, pp.367-375.
- PETOEFI, J.S. (ed) (1979)
Text vs Sentence: Basic Questions of Text Linguistics. Hamburg: Buske,
1979, 2 Vols. (= Papers in Textlinguistics 20, 1/2).
- RIESBECK, C.K. (1982)
Realistic Language Comprehension. In: W.G. Lehnert / M.H. Ringle
(eds): Strategies for Natural Language Processing. Hillsdale/NJ:
Erlbaum, 1982, pp.37-54.
- ROESNER, D.F. / J.H. LAUBSCH (1982)
Formalization of Argumentation Structures in Newspaper Texts. In:
COLING 82: Proceedings of the 9th International Conference on
Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia,
1982, pp.325-330.
- ROLLINGER, C.-R. (1984)
Zur Repräsentation der argumentativen Struktur von Texten. In: C.-R.
Rollinger (ed): Probleme des (Text-)Verstehens: Ansätze der Künstlichen
Intelligenz. Tübingen: Niemeyr, 1984, pp.107-124 (= Sprache und
Information 10).
- ROSENBERG, S.T. (1980)
Frame-Based Text Processing. In: D. Metzger (ed): Frame Conceptions
and Text Understanding. Berlin etc.: de Gruyter, 1980, pp.96-119
(= Research in Text Theory 5).
- ROSENBERG, S. / B. ROBERTS (1979)
Coreference in a Frame Database. In: IJCAI-79: Proceedings of the 6th
International Joint Conference on Artificial Intelligence. Tokyo,
August 20-23, 1979. Vol.2. 1979, pp.729-734.
- RUMELHART, D.E. (1975)
Notes on a Schema for Stories. In: D.G. Bobrow / A. Collins (eds):
Representation and Understanding: Studies in Cognitive Science. New
York etc.: Academic Pr., 1975, pp.211-236.
- SALLIS, P.J. (1978)
A Partial-Parsing Algorithm for Natural Language Text Using a Simple
Grammar for Arguments. In: ALLC Bulletin 6. 1978, pp.170-176.

- SCHANK, R.C. / R.P. ABELSON (1977)
Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures. Hillsdale/NJ: Erlbaum, 1977.
- SCHANK, R.C. / M. LEBOWITZ / L. BIRNBAUM (1980)
An Integrated Understander. In: American Journal of Computational Linguistics 6. 1980, No.1, pp.13-30.
- SMALL, S. (1980)
Word Expert Parsing: a Theory of Distributed Word-Based Natural Language Understanding. College Park/MD: Dept. of Computer Science, Univ. of Maryland, Sept. 1980 (= Computer Science Technical Report Series TR-954).
- THIBADEAU, R. / M.A. JUST / P.A. CARPENTER (1982)
A Model of the Time Course and Content of Reading. In: Cognitive Science 6. 1982, pp.157-203.
- WILENSKY, R. (1982)
Points: a Theory of the Structure of Stories in Memory. In: W.G. Lehnert / M.H. Ringle (eds): Strategies for Natural Language Processing. Hillsdale/NJ etc.: Erlbaum, 1982, pp.345-374.
- WINTER, E. (1978)
A Look at the Role of Certain Words in Information Structure. In: Informatics 3: Proceedings of a Conference held by the Aslib Co-ordinate Indexing Group. 2-4 April 1975, Emmanuel College, Cambridge. London: Aslib, 1978, pp.85-97.

4.2.1.2 Textrepräsentation

a) Speicherung mehrerer Textrepräsentationen in einer Wissensbasis

al) Entwicklung von inhaltsorientierten Auswahlmechanismen auf Textrepräsentationen

Die Aufgabe, ein vorgegebenes Wissensnetz mit einem anderen, größeren Wissensnetz zu vergleichen und festzustellen, ob es darauf paßt, entspricht dem Problem, einen isomorphen Teilgraphen zu ermitteln. Es gibt in der Graphentheorie dazu Algorithmen (z.B. LINGAS 1981), die alle eine hohe Zeitkomplexität aufweisen (NP-vollständig). Die dort erarbeiteten Verfahren bleiben jedoch weitgehend irrelevant für den Vergleich von Wissensnetzen, weil diese markierte Knoten und Kanten besitzen, so daß die Suche eines passenden Teilnetzes erheblich beschleunigt werden kann.

Zur Ermittlung der Ähnlichkeit zweier Wissensnetze können aber auch Verfahren des Pattern Matching herangezogen werden, die nicht nur Isomorphie, sondern auch partielle Übereinstimmungen feststellen, sowie unter mehreren zu vergleichenden Netzen diejenigen mit den besten Übereinstimmungen herausfinden. Eine allgemeine Einführung dazu gibt HAYES-ROTH 1978. Ein spezielles Verfahren ist in WINSTON 1980 dargestellt, das zwei Strukturbeschreibungen, z.B. von Geschichten, auf Ähnlichkeiten testet mit dem Ziel, Analogien herauszufinden. Dazu werden die Eigenschaften der Konzepte in den vorliegenden Beschreibungen sowie relationale Beziehungen zwischen ihnen herangezogen. Insbesondere wird durch Berücksichtigung einer Begriffshierarchie erreicht, daß Vergleiche nicht bloß auf Konzeptnamen stattfinden und damit bei Ungleichheit fehlschlagen, sondern

daß zwei Konzepte auch aufeinander passen, wenn zwischen ihnen die Ober-/Unterbegriffsrelation besteht.

Ein sehr einfacher Pattern Matcher, der Teil eines Simulationsmodells menschlichen Gedächtnisses ist, wird in ANDERSON/BOWER 1973 beschrieben. Wissen wird in diesem Modell durch Baumstrukturen repräsentiert. Erinnerung (also Retrieval) funktioniert über den Match eines Teilbaums, der eine Anfrage darstellt, und dem Wissensbaum. Es werden zu den terminalen Knoten des Teilbaums identische Knoten im Wissensbaum gesucht und anschließend von dort aus Kanten, die denen im Teilbaum gleichen. Ein Match erfolgt, wenn zu den Wegen im Teilbaum gleiche Wege im Wissensnetz gefunden wurden, d.h. die Anzahl der Kanten, ihre Markierung und Reihenfolge gleich sind. Damit realisiert dieser Algorithmus ein sehr rigides Matchverfahren.

Im wesentlich flexibleren Ansatz in SIMMONS 1984 gibt es für das Matchen zweier Knoten in einem semantischen Netz mehr Spielraum. Falls die Namen zweier Knoten nicht gleich sind, passen sie trotzdem aufeinander, wenn zwischen ihnen entweder Wege über Folgen von Ober-/Unterbegriffs- und Äquivalenzkanten oder über Folgen von Teile-, Lokalitäts- und UND-/ODER-Kanten existieren. Die Wegelängen sind jedoch beschränkt, um die semantische Nähe der Knoten zu gewährleisten.

MYLOPOULOS ET AL. 1975 beschreibt ein Verfahren, mit dem ein neues Wissensnetz einem existierenden Gesamtnetz hinzugefügt werden kann, so daß es in Abhängigkeit von den bestehenden Überlappungen beider Netze an den spezifischsten Stellen eingegliedert wird. Ein semantisches Netz ist in dieser Arbeit in Szenarios partitioniert, die über eine Spezialisierungshierarchie miteinander verknüpft sind. Das neue Netz wird denjenigen Szenarios hinzugefügt, die am tiefsten in dieser Hierarchie stehen und wo noch ein (partieller) Match zustande kommt. Die nicht gematchten Knoten werden hinzugefügt. Ähnlich wie in SIMMONS 1984 ist also auch hier ein Match auf verschiedenen Stufen der Begriffshierarchie möglich.

Das gleiche gilt für NISHIDA/DOSHITA 1979, wo die Anfrage einer Wissensbasis, die aus einer Menge logischer Formeln besteht, dargestellt wird. Eine Frage ist als ein Pattern repräsentiert, das mit den Formeln der Wissensbasis, die ebenfalls als Pattern interpretiert werden, verglichen wird. Über eine Indexstruktur auf den Formelsymbolen werden Vergleichskandidaten ausgewählt. Auch hier ist ein erfolgreicher Vergleich über den Fall der bloßen Identität hinaus möglich durch Einbeziehung der Begriffshierarchie und durch Berücksichtigung von Paraphrasen gleicher Sachverhalte. Mit einem Pattern-Matching-Verfahren auf einer Wissensbasis, die aus logischen Formeln besteht, Retrievaloperationen zu realisieren, ist ungewöhnlich und wurde in dieser Arbeit zur Effizienzsteigerung gewählt.

Die übliche Vorgehensweise zur Anfrage von logischen Wissensbasen besteht in der Verwendung von Theorembeweisern. In GREEN 1969 ist ein Frage-Antwort-System auf dieser Basis beschrieben worden. Die Wissensbasis ist eine Ansammlung logischer Formeln, Fragen werden als Theoreme formuliert, und durch Resolution wird die Antwort generiert. Der Theorembeweiser-Ansatz für Frage-Antwort-Systeme wird für Retrievaloperationen auf semantischen Netzen dadurch relevant, daß semantische Netze als Notationsvariante einer logischen Repräsentation aufgefaßt werden können. Es wird z.B. in SCHUBERT 1975 eine Erweiterung semantischer Netze gegeben, so daß ihre Darstellungsmächtigkeit der der Prädikatenlogik äquivalent wird. HENDRIX 1979 erreicht das gleiche durch Partitionierung von Netzen und beschreibt außerdem einen Retrievalmechanismus, der als Anfrage ein semantisches Netz entgegennimmt, es als Theorem interpretiert,

auf der Wissensbasis beweist und damit die Antwort generiert. In DELIYANNI/KOWALSKI 1979 werden semantische Netze als Horn-Klausen repräsentiert, so daß ähnlich wie in GREEN 1969 Antworten durch Einsatz eines Resolutionsverfahrens generiert werden können.

KOLODNER 1984 stellt einen völlig anderen Ansatz vor. Dieser Arbeit liegt der Anspruch zugrunde, ein kognitiv adäquates Modell menschlichen Gedächtnisses und Erinnerns zu formulieren. Aufwendige Vergleichsoperationen zwischen Repräsentationsstrukturen, die mit dem Anwachsen der Wissensbasis komplexer werden und somit kognitiv nicht plausibel sind, werden durch eine entsprechend vorstrukturierte Wissensbasis vermieden. Alle Wissenseinheiten sind so dargestellt, daß ihre Unterschiede indexiert und ihre Gemeinsamkeiten durch Verallgemeinerungen zusammengefaßt werden. Sind in einer Anfrage Konzepte erwähnt, die in dieser Indexstruktur wiederzufinden sind, ist das Retrieval ein recht einfacher, zielgerichteter Vorgang. Ist dies nicht der Fall, müssen durch geeignete Inferenzregeln Indizes ausgewählt werden, die alternative Bereiche der Wissensbasis für eine Suche zugänglich machen.

In TOPIC II zu realisierende Retrievalfunktionen operieren auf einem semantischen Netz, dessen Knoten aus Frames bestehen. Die geschilderten Matching-Verfahren stellen somit wichtige Vorarbeiten dar, berücksichtigen aber nicht die Darstellung der Netzknoten als Frames, so daß hierzu neue Verfahren zu entwickeln sind. Über Faktenabfragen hinaus, wie sie von den bisherigen Ansätzen realisiert wurden, liegt in TOPIC II der Schwerpunkt auf der Bereitstellung von Operationen für Referenzretrieval.

a2) Berücksichtigung von Änderungen der für die Analyse benutzten Weltwissensbasis

Zur automatischen Modifikation einer Suchfrage gibt es unterschiedliche Motivationen und deshalb verschiedene Ansätze. In DEMOLOMBE 1981 werden bestimmte Typen syntaktisch unsinniger, in Logik formulierter Fragen in wohlgeformte Fragen umgewandelt, indem die darin vorkommenden ungebundenen Variablen auf die Wertebereiche der Prädikatargumente, die sie besetzen, restringiert werden. In HAMMER/ZDONIK 1980 werden Fragen so reformuliert, daß ihre neuen Auswahlbedingungen bei gleichbleibender Fragesemantik auf der zugrunde liegenden Wissensbasis effizienter auszuwerten sind. In STONEBRAKER 1975 werden aus Suchfragen, die virtuelle Relationen einer externen Sicht benutzen, unter Verwendung der Sichtdefinition Fragen auf Basisrelationen abgeleitet. Im Information Retrieval existieren Verfahren, die Indexterme einer Suchfrage in Abhängigkeit von der Relevanzbewertung des Suchergebnisses durch den Benutzer neu gewichten, so daß eine neue, exaktere Frage entsteht (SALTON/McGILL 1983). Schließlich gibt es noch Verfahren zur Fragemodifikation mit dem Ziel, Änderungen im Schema einer Datenbank oder Datenmodellwechsel zu kompensieren. Dieser letzte Bereich entspricht unmittelbar der in TOPIC II zu behandelnden Aufgabe, Änderungen im Weltwissen in den Retrievaloperationen zu berücksichtigen. Im folgenden werden deshalb diese Ansätze näher betrachtet.

In Anwendungsumgebungen immer wieder auftretende Änderungen im Schema einer Datenbank erfordern eine entsprechende Modifikation aller Anwendungsprogramme, die auf diese Datenbank zugreifen. Mit dem Ziel, diesen zeitaufwendigen und teuren Folgeaufwand durch automatische Verfahren zu unterstützen, wurde in SU 1976 ein Ansatz skizziert, wie - entsprechend den Änderungen im Schema - die Datenbankaufrufe in Anwendungsprogrammen automatisch zu konvertieren sind. Einige Konversionstypen erfordern dabei zusätzliche Modifikationen auf der Ebene der Wirtssprache, die wegen ihrer hohen Komplexität jedoch nur halbautomatisch vorzunehmen sind.

Ein Verfahren zur Konversion von Suchfragen in einem einfachen Netzwerkmodell ist in SHNEIDERMAN/THOMAS 1980 beschrieben. In diesem Modell werden Suchfragen durch ihre Wegebeschreibungen im Netz dargestellt. Eine Änderung im Netzwerk ist auf diese Weise einfach in entsprechende Änderungen der Wegebeschreibungen in den Suchfragen umzusetzen.

In einer späteren Arbeit der gleichen Autoren (SHNEIDERMAN/THOMAS 1982) werden fünfzehn Schematransformationen für das relationale Datenmodell beschrieben, die als besonders praxisrelevant eingeschätzt werden. Die Transformationen sind in verschiedene Typen unterteilt. Informationserhaltende Transformationen (z.B. Hinzufügen eines Attributs) verursachen keinen Informationsverlust in der Datenbank. Datenabhängige Transformationen (z.B. Löschen eines Attributs in einem Schlüssel) erfordern eine Überprüfung ihrer Zulässigkeit im neuen Schema, während die Zulässigkeit programmabhängiger Transformationen (z.B. Löschen eines in einer Suchfrage benutzten Attributs) durch Inspektion der Anwendungsprogramme geprüft werden muß. Erweisen sich in einem Anwendungsfall alle Schematransformationen als zulässig, so sind die Anwendungsprogramme automatisch transformierbar und sind bedeutungsäquivalent zu ihren ursprünglichen Versionen. Ein manuelles Eingreifen des Datenbankadministrators wird nur notwendig, wenn festgestellt wird, daß bedeutungserhaltende Transformationen nicht möglich sind, weil es Konsistenzverletzungen bei datenabhängigen Transformationen gibt bzw. die Logik eines Anwendungsprogramms dies bei programmabhängigen Transformationen nicht zuläßt.

In SU/LAM/LO 1981 wird ein semantisches Datenmodell herangezogen, dessen semantische Relationen es ermöglichen, Teile der sonst in Anwendungsprogrammen verborgenen Semantik (z.B. Datenabhängigkeiten, Integritätsbedingungen) explizit im Schema darzustellen. Unter Benutzung dieser Semantikbeschreibungen werden Programme in Zugriffspfadgraphen übersetzt. Auf dieser Darstellungsebene findet die eigentliche Programmtransformation statt, der resultierende Graph wird anschließend in ein Programm zurückübersetzt. Alle in dieser Arbeit behandelten Transformationen sind bedeutungserhaltend; es wird ferner nur der Datenbankinteraktionsteil der Programme betrachtet, nicht dagegen notwendige Änderungen auf der Ebene der Wirtssprache.

KATZ/GOODMAN 1983 stellt einen Ansatz vor, Anfragen auf einem externen, globalen Schema, das mehrere lokale Schemata in sich vereint, in Anfragen auf diesen lokalen Schemata zu überführen. Es wird dazu der in STONEBRAKER 1975 beschriebene Ansatz herangezogen. Interessant in bezug auf Fragetransformationen für Schemaänderungen wird diese Arbeit dadurch, daß über die einfache Abbildung virtueller Relationen auf Basisrelationen hinaus der Verwendung verschiedener Begriffe zwischen den Schemata Rechnung getragen wird. Als Beziehungen zwischen den verschiedenen Begriffen sind Synonymie, Ober-/Unterbegriffsrelation und Wechsel in den Maßeinheiten berücksichtigt. Auch hier sind alle Fragetransformationen bedeutungserhaltend.

Die Abbildung von Operationen in einem externen relationalen Schema auf ein konzeptuelles Schema in einem anderen Datenmodell wird in HWANG/DAYAL 1983 dargestellt. Als Zwischenstufe in dieser Abbildung wird ein dem konzeptuellen Schema äquivalentes Schema, das auf dem Entity-Relationship-Modell basiert, herangezogen. Operationen im relationalen Schema werden zunächst in Operationen auf dem Entity-Relationship-Schema überführt und von dort aus in Operationen auf dem konzeptuellen Schema übersetzt. Alle Transformationen sind bedeutungserhaltend.

Eine allgemeine Behandlung der Eigenschaften von Schema- und Fragetransformationen zwischen Datenmodellen auf formaler Ebene ist in KALINICHENKO 1978 zu finden.

Alle hier vorgestellten Ansätze und Verfahren beziehen sich entweder ausschließlich auf bedeutungserhaltende Transformationen oder greifen auf manuelle Interaktion zurück. In den in TOPIC II anvisierten Fragemodifikationen zur Berücksichtigung von Weltwissensänderungen sind dagegen auch Transformationen systemintern zu behandeln, die nicht bedeutungserhaltend sind. Die hierdurch zwangsläufig auftretenden Fehler und Ungenauigkeiten sind jedoch im Gegensatz zu den den bisherigen Verfahren zugrunde liegenden Anwendungsbereichen vertretbar, da die Suchoperationen hier in erster Linie Referenzfunktion haben und der Benutzer Kontrollmöglichkeit über die Korrektheit der Suchergebnisse hat, zumal er vom System eine entsprechende Warnung erhält, wenn er mit Retrievalfehlern rechnen muß. Ein weiterer neuer Aspekt kommt durch die Behandlung anderer Typen von Repräsentationsstrukturen hinzu.

b) Wissensextraktion aus Texten:

Erweiterung der verfügbaren Wissensrepräsentationsstrukturen

Die strukturelle Ähnlichkeit der Repräsentation von Sichten und Versionen (wenn auch ihre Semantik verschieden ist) erlaubt im folgenden die Betrachtung dafür relevanter Vorarbeiten unter einem gemeinsamen Blickwinkel.

Zur gleichzeitigen Darstellung verschiedener Sichten auf ein Objekt (z.B. ein Mikrorechner als Computer und als intelligentes Terminal) bietet die Wissensrepräsentationssprache KRL (BOBROW/WINOGRAD 1977) den Mechanismus von "perspectives" an, mit denen ein Objekt gleichzeitig als Instanz verschiedener Prototypen dargestellt werden kann. Dieses Sprachkonstrukt ist ausschließlich syntaktisch definiert. Es gibt keine Möglichkeit, zumindest die Semantik in einer konkreten Anwendung von KRL über Integritätsbedingungen zu spezifizieren, wenn schon die Semantik des Konstrukts an sich unklar bleibt. KL-ONE (BRACHMAN ET AL. 1979) stellt ein Konstrukt zur Verfügung, das einer Repräsentation eines Objekts zusätzliche Rollen ("Roles") zuweist. Es wird dabei unterschieden zwischen generischen Rollenindikatoren für Prototypen ("RoleSets") und ihren Ausprägungen ("IRoles"), die auf Instanzen definiert sind. Im Unterschied zu KRL können Integritätsbedingungen für eine Rollenzuordnung spezifiziert werden. Die für eine Rollenbesetzung in Frage kommenden Objekte sind so auf bestimmte Objektklassen einschränkbar (z.B. muß der Präsident einer Organisation Angestellter dieser Organisation sein); die Anzahl an Ausprägungen eines generischen Rollenindikators ist spezifizierbar und seine Modalität kann angegeben werden (z.B. optional, obligatorisch).

Im Bereich von Wissensrepräsentationssprachen der Künstlichen Intelligenz stellt nur KL-ONE Mechanismen zur Verfügung, die Semantik der Repräsentation verschiedener Sichten, bezogen auf eine Anwendung, darzustellen. Es fehlen grundsätzlich aber auch hier weitergehende Möglichkeiten, zu erfüllende Beziehungen zwischen verschiedenen Sichten eines Objekts, wie z.B. Identität, Überlappung oder andere Arten von Abhängigkeiten zwischen Objekteigenschaften, darzustellen.

In diesem Punkt stärker ins Detail gehende Arbeiten finden sich in der Datenbankliteratur, die Abhängigkeiten zwischen externen Schemata und ihrem zugehörigen konzeptuellen Schema behandeln. Ein externes Schema ist

durch eine Serie von Operationen aus einem konzeptuellen Schema ableitbar und stellt eine ganz bestimmte Sicht auf die zugrunde liegende Datenbank dar. Verschiedene externe Schemata sind dann unterschiedliche Sichten auf ein und dieselbe Datenbank (z.B. SEVCIK/FURTADO 1978). Änderungsoperationen in einem externen Schema sind abzubilden auf das konzeptuelle Schema, weil nur das konzeptuelle Schema physisch repräsentiert ist, während das externe Schema lediglich definitiv vorhanden sind, d.h. aus dem konzeptuellen Schema berechnet wird. Die Ausführung einer Änderungsoperation in einem externen Schema kann somit Seiteneffekte auf das konzeptuelle Schema haben, die wiederum (möglicherweise ungewollte) Seiteneffekte in anderen externen Schemata haben können. Abbildungen zwischen externem und konzeptuellem Schema, die keine oder möglichst wenig Seiteneffekte haben, sind in einigen Arbeiten untersucht worden. Gleichzeitig werden damit Beziehungen zwischen verschiedenen externen Schemata untersucht: PAOLINI/PELEGATTI 1977, FURTADO/SEVCIK/dosSANTOS 1979, SPYRATOS 1980, BANCILHON/SPYRATOS 1981. Eine Änderungsoperation auf einem externen Schema kann schließlich sogar Seiteneffekte im gleichen Schema haben (KELLER 1982). Alle diese Ansätze sind zwar für die eigenen Arbeitsziele interessant, bieten aber keinen unmittelbaren Bezug, weil sie nicht unmittelbar konzeptorientiert sind, sondern Abhängigkeiten über Operationsfolgen untersuchen.

Eine den eigenen geplanten Arbeiten viel näher kommende Behandlung von Sichten und Versionen und ihrer Beziehungen zueinander ist in NEUMANN/HORNUNG 1982 zu finden. Im Rahmen einer Datenbank für CAD-Anwendungen wird die Darstellung verschiedener Sichten sowie unterschiedlicher Entwurfsalternativen (Versionen) eines Objekts beschrieben. Abhängigkeiten zwischen ihnen werden in einem Abhängigkeitsgraphen dargestellt, auf dem Konsistenzabhängigkeiten definiert sind (abhängig inkonsistent, unabhängig inkonsistent). Die Zulässigkeit verschiedener Änderungsoperationen ist von dem aktuellen Konsistenzzustand abhängig.

L i t e r a t u r v e r z e i c h n i s

- ANDERSON, J. / G. BOWER (1973)
Human Associative Memory. Winston-Wiley, 1973.
- BANCILHON, F. / N. SPYRATOS (1981)
Update Semantics of Relational Views. In: ACM Transactions on Database Systems 6. 1981, No.4, pp.557-575.
- BOBORW, D.G. / T. WINOGRAD (1977)
An Overview of KRL, a Knowledge Representation Language. In: Cognitive Science 1. 1977, No.1, pp.3-46.
- BRACHMAN, R.J. / D.G. BOBROW / P.R. COHEN / J.W. KLOVSTAD / B.L. WEBER / W.A. WOODS (1979)
An Introduction to KL-ONE. In: Research on Knowledge Representation for Natural Language Understanding. BBN Report No.4274, 1979.
- DELIYANNI, A. / R.A. KOWALSKI (1979)
Logic and Semantic Networks. In: Communications of the ACM 22. 1979, No.3, pp.184-192.

- DEMOLOMBE, R. (1981)
Assigning Meaning to Ill-Defined Queries Expressed in Predicate Calculus Language. In: H. Gallaire, J. Minker, J.M. Nicolas (eds): Advances in Data Base Theory, Vol.1. Plenum Press, 1981, pp.367-395.
- FURTADO, A.L. / K.C. SEVCIK / C.S. dos SANTOS (1979)
Permitting Updates Through Views of Data Bases. In: Information Systems 4. 1979, pp.269-283.
- GREEN, C.C. (1969)
Theorem Proving by Resolution as a Basis for Question Answering Systems. In: B. Meltzer, D. Michie (eds): Machine Intelligence, Vol.4. American Elsevier, 1969, pp.183-208.
- HAMMER, M. / S.B. ZDONIK (1980)
Knowledge-Based Query Processing. In: Proc. Int. Conf. on Very Large Data Bases, 1980, pp.137-147.
- HAYES-ROTH, F. (1978)
The Role of Partial and Best Matches in Knowledge Systems. In: D.A. Waterman, F. Hayes-Roth (eds): Pattern-Directed Inference Systems. Academic Press, 1978, pp.557-574.
- HENDRIX, G. (1979)
Encoding Knowledge in Partitioned Networks. In: N.V. Findler (ed): Associative Networks. Academic Press, 1979, pp.51-92.
- HWANG, H.-Y. / U. DAYAL (1983)
Using the Entity-Relationship Model for Implementing Multi-Model Database Systems. In: P.P. Chen (ed): Entity-Relationship Approach to Information Modeling and Analysis. Amsterdam etc.: North-Holland, 1983, pp.237-258.
- KALINICHENKO, L.A. (1978)
Data Models Transformation Method Based on Axiomatic Data Model Extension. In: Proc. Int. Conf. on Very Large Data Bases, 1978, pp.549-555.
- KATZ, R.H. / N. GOODMAN (1983)
View Processing in MULTIBASE, a Heterogeneous Database System. In: P.P. Chen (ed): Entity-Relationship Approach to Information Modeling and Analysis. Amsterdam etc.: North-Holland, 1983, pp.259-279.
- KELLER, A.M. (1982)
Updates to Relational Databases Through Views Involving Joins. In: P. Scheuermann (ed): Improving Database Usability and Responsiveness. Academic Press, 1982, pp.363-384.
- KOLODNER, J.L. (1984)
Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model. Hillsdale/NJ: Erlbaum, 1984.
- LINGAS, A. (1981)
Certain Algorithms for Subgraph Isomorphism Problems. In: E. Astesiano, C. Böhm (eds): Proc. of 6th Colloquium on Trees in Algebra and Programming, 1981. Springer-Verlag, pp.290-307.
- MYLOPOULOS, J. / P. COHEN / A. BORGIDA / L. SUGAR (1975)
Semantic Networks and the Generation of Context. In: Proc. Int. Joint Conf. on Artificial Intelligence, 1975, pp.134-142.

- NEUMANN, T. / C. HORNING (1982)
Consistency and Transactions in CAD Database. In: Proc. Int. Conf. on Very Large Data Bases, 1982, pp.181-188.
- NISHIDA, T. / S. DOSHITA (1979)
The Framework of Knowledge Representation and its Retrieval in LGS - the Literature Guide System. In: Proc. Int. Joint Conf. on Artificial Intelligence, 1979, pp.662-664.
- PAOLINI, P. / G. PELEGATTI (1977)
Formal Definition of Mappings in a Data Base. In: Proc. Int. Conf. on Management of Data, 1977, pp.40-46.
- SALTON, G. / M.J. MCGILL (1983)
Introduction to Modern Information Retrieval. New York etc.: McGraw-Hill, 1983.
- SCHUBERT, L. (1975)
Extending the Expressive Power of Semantic Networks. In: Proc. Int. Joint Conf. on Artificial Intelligence, 1975, pp.158-164.
- SEVCIK, K.C. / A.L. FURTADO (1978)
Complete and Compatible Sets of Update Operations. In: Proc. Int. Conf. on Data Base Management Systems, 1978, pp.247-260.
- SHNEIDERMAN, B. / G. THOMAS (1980)
Path Expressions for Complex Queries and Automatic Database Program Conversion. In: Proc. Int. Conf. on Very Large Data Bases, 1980, pp.33-44.
- SHNEIDERMAN, B. / G. THOMAS (1982)
An Architecture for Automatic Relational Database System Conversion. In: ACM Transactions on Database Systems 7. 1982, No.2, pp.235-257.
- SIMMONS, R.F. (1984)
The Costs of Inheritance in Semantic Networks. In: Proc. 10th Int. Conf. on Computational Linguistics, 1984, pp.71-74.
- SPYRATOS, N. (1980)
Translation Structures of Relational Views. In: Proc. Int. Conf. on Very Large Data Bases, 1980, pp.411-416.
- STONEBRAKER, M. (1975)
Implementation of Integrity Constraints and Views by Query Modification. In: Proc. Int. Conf. on Management of Data, 1975, pp.65-78.
- SU, S.Y.W. (1976)
Application Program Conversion Due to Data Base Changes. In: P.C. Lockemann, E.J. Neuhold (eds): Systems for Large Data Bases. Amsterdam etc.: North-Holland, 1976, pp.143-157.
- SU, S.Y.W. / H. LAM / D.H. LO (1981)
Transformation of Data Traversals and Operations in Application Programs to Account for Semantic Changes of Databases. In: ACM Transactions on Database Systems 6. 1981, No.2, pp.255-294.

WINSTON, P.H. (1980)

Learning and Reasoning by Analogy. In: Communications of the ACM 23. 1980, No.12, pp.689-703.

4.2.1.3 Textkondensierung

Im Rahmen der wissensbasierten Textkondensierung kann auf mehrere unterschiedliche Konzeption für die Realisierung von Textzusammenfassungen zurückgegriffen werden:

- Mikro/Makropropositionen-Modell
- Story Grammars
- Konnektivitätsansatz
- Inferenz-orientierter Ansatz

Das Textzusammenfassungskonzept der Mikro/Makropropositionen beruht auf sog. Makroregeln. Dabei handelt es sich um Tilgungs-, Generalisierungs- und Rekonstruktionsregeln, die auf den in einer Textwissensbasis verfügbaren einzelnen Propositionen eines Texts (Mikropropositionen) operieren, um semantische Information zu reduzieren, d.h. Makropropositionen zu erzeugen. Eine Makroproposition ist logisch in den ihr zugeordneten Mikropropositionen enthalten. Entsprechend kann diese Beziehung in einem Textgraphen auch als propositionale Spezialisierungsrelation interpretiert werden derart, daß hierarchisch "hohe" Knoten (Makropropositionen) Zusammenfassungen des propositionalen Gehalts mit diesem Knoten verbundener hierarchisch "niedrigerer" Knoten (Mikropropositionen) sind (DIJK 1980). Das Mikro/Makropropositionenmodell beansprucht kognitionspsychologische Plausibilität und ist durch umfangreiche experimentelle Begleitforschung auch empirisch validiert (KINTSCH/DIJK 1978, DIJK 1979). Die Übertragung der diskurstheoretischen Diskussion auf die Zusammenfassung von Fachtexten (DIJK 1977) wird im Rahmen des SUSY-Abstracting-Systems auch in Form einer Implementationslösung für dieses Konzept verfolgt (FUM et al. 1982). Einen bereits sehr entwickelten konstruktiven Versuch dokumentiert CORREIRA (1980), dessen Zusammenfassungsalgorithmus nicht allein Makropropositionen auf einer festgelegten hierarchischen Ebene traversiert, sondern aktuelle Makropropositionen an bereits verarbeitete adaptiert.

Im Mittelpunkt des Textzusammenfassungskonzepts von 'story grammars' stehen für Kondensate obligatorische und fakultative Komponenten des nichtterminalen Vokabulars von Story-Produktionsregeln (RUMELHART 1975). Dieses Reduktionskonzept geht über das der Mikro/Makropropositionen hinaus, da die Beziehung einer strengen logischen Implikation zwischen der ursprünglichen Story- und der reduzierten Kondensatstruktur wegen der zusätzlichen Elimination ganzer Teilbäume nicht mehr gewährleistet ist. Auch diese Konzeption zur Informationsreduktion von Texten ist im Rahmen kognitionspsychologischer Experimente überprüft worden (RUMELHART 1977). In einer Rumpfform findet sich dieses Konzept der Ausdünnung von Repräsentationsstrukturen etwa in deklarativer Form in den 'sketchy scripts' des FRUMP-Systems (DeJONG 1979) bzw. in prozeduraler Form in diversen Listentilgungsfunktionen, mit denen instantiierte Textschemata zurückgeschnitten werden (SIMMONS 1984).

Die Zusammensetzung atomarer Wissensrepräsentationskonstrukte zu komplexen Textrepräsentationsstrukturen graphentheoretisch zu interpretieren und maximal konnektive Subgraphen bzw. Aktivierungscluster zu isolieren, die den (die) thematischen Schwerpunkt(e) eines Texts repräsentieren,

charakterisiert den Konnektivitätsansatz zur Textkondensierung (TAYLOR 1974, LEHNERT 1982). Seine technische Realisierung im Rahmen von Textverstehenssystemen erscheint weniger problematisch - gesetzt, die Wissensrepräsentationsform des jeweiligen Systems ist graphentheoretisch interpretierbar - als etwa das Mikro/Makropropositionen-Modell, das die Erkennung von Mikropropositionen des Textes voraussetzt. LEHNERT et al. (1981) zeigen darüber hinaus, daß ihr auf "plot units" basierendes Konnektivitätsmodell der Textzusammenfassung höhere empirische Adäquatheit besitzt als das entsprechende Konzept der "story grammars".

Beim inferenz-orientierten Ansatz werden alle aus Hintergrundwissen oder anderen voraussetzbaren Wissensquellen inferierbaren Informationen in Textzusammenfassungen unterdrückt und lediglich die "neue" Information des Texts vermittelt (COOK et al. 1984). Damit kehrt sich die Reduktionsproblematik für Textzusammenfassungen um in ein Textgenerierungsproblem.

Dieses letzte Konzept kontrastiert am meisten mit einem der wenigen Versuche, ein informationswissenschaftliches Textkondensierungsmodell (oberhalb der Ebene von handhabungsorientierten "abstracting guidelines") zu entwickeln (HUTCHINS 1977). Es geht mit seinem "aboutness"-Konzept aber auch von den Reduktionskonzepten des Mikro/Makropropositionen- und "story grammar"-Modells ab. Hintergrund ist das idealisierte Konstrukt einer verallgemeinerten Makrostruktur, d.i. Wissen, das alle potentiellen Benutzer eines Informationssystems teilen (und nur aufgrund dessen sie erfolgreich Suchfragen spezifizieren können). Die "aboutness" eines Dokuments in Form eines Kondensats zu beschreiben, ist dann identisch mit dem Problem, den Ausschnitt der allgemeinen Makrostruktur einzugrenzen, den der jeweilige Text abdeckt. Damit ist der Übereinstimmungsbereich zwischen einem spezifischen Dokumentinhalt und dem generalisierten Wissen der Benutzer eines Informationssystems fixiert, der einen einzelnen Benutzer dann erst in die Lage versetzt, die Relevanz eines Dokuments aufgrund eines ihm zugeordneten Kondensats einzuschätzen.

L i t e r a t u r v e r z e i c h n i s

- COOK, M.E. / W.G. LEHNERT / D.D. McDONALD (1984)
Conveying Implicit Content in Narrative Summaries. In: COLING 84: Proceedings of the 10th International Conference on Computational Linguistics & 22nd Annual Meeting of the Association for Computational Linguistics. 2-6 July 1984, Stanford University, CA. o.O.: ACL 1984, pp.5-7.
- CORREIRA, A. (1980)
Computing Story Trees. In: American Journal of Computational Linguistics 6. 1980, Nos.3-4, pp.135-149.
- DeJONG, G.F. (1979)
Skimming Stories in Real Time: an Experiment in Integrated Understanding. Yale University, May 1979. (Ph.D.Thesis)
- DIJK, T.A. van (1977)
Perspective Paper: Complex Semantic Information Processing. In: D.E. Walker / H. Karlgren / M. Kay (eds): Natural Language in Information Science: Perspectives and Directions for Research. Results of a Workshop on Linguistics and Information Science. Biskops-Arne, Sweden, 3-5 May 1976. Stockholm: Skriptor, 1977, pp.127-163 (= FID Publication 551).

- DIJK, T.A. van (1979)
Recalling and Summarizing Complex Discourse. In: W. Burghardt / K. Hölker (eds): Text Processing: Papers in Text Analysis and Text Description. Berlin etc.: W. de Gruyter, 1979, pp.49-118 (= Research in Text Theory 3).
- DIJK, T.A. van (1980)
Macrostructures: an Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition. Hillsdale/NJ: Erlbaum, 1980.
- FUM, D. / G. GUIDA / C. TASSO (1982)
Forward and Backward Reasoning in Automatic Abstracting. In: COLING 82: Proceedings of the 9th International Conference on Computational Linguistics. Prague, July 5-10, 1982. Prague: Academia, 1982, pp.83-88.
- HUTCHINS, W.J. (1977)
On the Problem of "Aboutness" in Document Analysis. In: Journal of Informatics 1. 1977, No.1, pp.17-35.
- KINTSCH, W. / T.A. van DIJK (1978)
Toward a Model of Text Comprehension and Production. In: Psychological Review 85. 1978, No.5, pp.363-394.
- LEHNERT, W.G. (1982)
Plot Units: a Narrative Summarization Strategy. In: W.G. Lehnert / M.H. Ringle (eds): Strategies for Natural Language Processing. Hillsdale/NJ etc.: Erlbaum, 1982, pp.375-412.
- LEHNERT, W.G. / J.B. BLACK / B.J. REISER (1981)
Summarizing Narratives. In: IJCAI-81: Proceedings of the 7th International Joint Conference on Artificial Intelligence. 24-28 August 1981, University of British Columbia, Vancouver/BC, Canada. Vol.1. Menlo Park/CA: AAI, 1981, pp.184-189.
- RUMELHART, D.E. (1975)
Notes on a Schema for Stories. In: D.G. Bobrow / A. Collins (eds): Representation and Understanding: Studies in Cognitive Science. New York etc.: Academic Press, 1975, pp.211-236.
- RUMELHART, D. E. (1977)
Understanding and Summarizing Brief Stories. In: D. LaBerge / S.J. Samuels (eds): Basic Processes in Reading: Perception and Comprehension. Hillsdale/NJ: Erlbaum, 1977, pp.265-303.
- SIMMONS, R.F. (1984)
Computations from the English: a Procedural Logic Approach for Representing and Understanding English Texts. Englewood Cliffs/NJ: Prentice-Hall, 1984, pp.174-193 (= Ch.9).
- TAYLOR, S.L. (1974)
Automatic Abstracting by Applying Graphical Techniques to Semantic Networks. Evanston/Illinois: Northwestern University, Computer Science, (Aug) 1974. (Ph.D.Thesis)

4.2.2 Textinformationsmanagement in TOPOGRAPHIC

4.2.2.1 Formen graphischer Interaktion

4.2.2.1.1 Benutzerschnittstellen

Mit fallenden Hardwarepreisen werden seit Mitte der Siebziger Jahre Computer zunehmend auch ungeübten Benutzern zugänglich, womit sich automatisch das Problem stellt, "benutzerfreundliche" Schnittstellen zu schaffen. Die Kriterien für die Konstruktion solcher Systeme werden häufig aus eigener mißlicher Erfahrung abgeleitet (NIEVERGELT 1982) und führen zu pragmatischen Ansätzen, wie dem Konzept der "Sites, Modes and Trails" (NIEVERGELT 1979, NIEVERGELT/WEYDERT 1982), dessen Grundidee darin liegt, den Benutzer jederzeit über den Systemzustand zu informieren.

Daraus läßt sich die Forderung nach übersichtlich strukturierten Kommandostrukturen ableiten. Anhand von zwei Beispielen sollen die Möglichkeiten graphisch-interaktiver Kommandosprachen aufgezeigt werden. Im System OMEGA (POWELL/LINTON 1983) wurde in einem eingegrenzten Aufgabenbereich (Programmerstellung) mit Hilfe von Piktogrammen das Prinzip "what you see is what you get" realisiert. Das interaktive Betriebssystem XS2 (STELOVSKY 1984) stellt eine applikationsunabhängige Kommandosprache zur Verfügung, die für gleich strukturierte Daten auch die gleichen Kommandos bereithält (s.a. NIEVERGELT 1979). Die baumartigen Kommando- und Datenstrukturen werden graphisch präsentiert und können mit einer Maus angewählt werden. Die Verwendung von Piktogrammen zur Konstruktion kognitiv günstiger Kommandosprachen wird in STAUFER 1984 diskutiert.

Sowohl Informatik, wie auch Kognitionswissenschaft arbeiten zur Zeit daran, daß Problemfeld Mensch-Maschine-Kommunikation zu systematisieren (DEHNING et al. 1981, JAGODZINSKI 1983) und die kognitive Adäquanz verschiedener Kommunikationsformen empirisch zu erarbeiten (WAERN/ROLLENHAGEN 1983). Daraus ergibt sich die Forderung nach Repräsentationsverfahren für Dialoge (JAGODZINSKY 1983). Unter diesen Techniken markieren einerseits die Übergangsnetzwerke (OBERQUELLE 1981, DEHNING et al. 1981), die auch zur Modellierung allgemeiner Schnittstellenprobleme eingesetzt werden (GMD 1982), und andererseits die in der künstlichen Intelligenz entwickelten Wissensrepräsentationsverfahren die gegensätzlichen Positionen. Letztere werden sowohl im Sinne einer aufgaben-zentrierten (POLLITT 1981) als auch einer benutzerorientierten (MORIK 1984) Dialogmodellierung eingesetzt.

4.2.2.1.2 Präsentation von Netzwerken

Der komplexe Aufbau netzwerkartig strukturierter Daten, wie sie in zunehmendem Maß in Wissensrepräsentation und Datenbanktechnik eingesetzt werden, läßt eine graphisch interaktive Präsentation und Manipulation dieser Strukturen als wünschenswert erscheinen (CAHN 1980, ZDYBEL et al. 1981). Im Gegensatz zur Präsentation von Netzwerken, die topographisch weitestgehend festgelegt sind, wie z.B. Landkarten, stellt sich schon bei relativ einfachen Datenstrukturen, wie z.B. Bäumen, die Frage nach der einsichtigsten Darstellung (LEWIS 1975), die nicht nur der Struktur, sondern auch der Semantik der Daten gerecht wird. Eine Vielzahl von Heuristiken zur Darstellung von allgemeinen Netzen, in die sowohl die Netzstruktur (Vernetzungsgrad, Umfang) wie auch die Präsentationsintention eingehen, werden in BERTIN 1982 erarbeitet. Einige dieser Ergebnisse, die schon in einer früheren Veröffentlichung vorgestellt wurden, greift OBERQUELLE 1981 auf und entwickelt sie weiter. Anhand der so gewonnenen

Kriterien werden dann aus der Literatur entnommene Netzdarstellungen überarbeitet. Diese Darstellungen zeigen sowohl die Stärken, wie auch die Schwächen dieser Regeln. Die überarbeiteten Graphiken sind durchweg übersichtlicher geworden, aber zwischen den explizit erwähnten Designregeln und den manuell erstellten Abbildungen klafft noch eine konzeptionelle Lücke, die vor der Implementation dieser Regeln noch zu schließen wäre.

Da es zur "Animation" von Navigationsvorgängen in Wissensnetzen keine Vorarbeiten gibt, sollen aus den zahlreichen Verfahren der Dynamic Graphics hier nur diejenigen herausgegriffen werden, die geeignet erscheinen, die Präsentation von Wissenstrukturen zu unterstützen. Als weitere Einschränkung kommt hinzu, daß der "Film" nicht vorproduziert werden kann, sondern während des Dialogs situationsspezifisch generiert werden muß. Damit scheiden alle Verfahren der Key-Frame-Animation aus, ebenso eine nicht realzeitbezogene Direct Animation. Diese Kriterien gelten auch für die Laufzeitanimation von Programmen, weshalb einige Arbeiten zu diesem Thema untersucht werden sollen. Es ergibt sich, daß die Unterstützung der Programm-Animation entweder auf bescheidene Stilmittel beschränkt ist, wie wandernde Markierungen und Beleuchten von Teilstrukturen (SEGLER 1984), oder nicht datengesteuert ist, so daß der zum Programm gehörige "Film" letztlich vom Programmierer durch "Regieanweisungen" festgelegt werden muß (BAECKER 1975). Eine Zwischenform stellt das System Antics (DIONNE/MACKWORTH 1978) dar, das zu Ausbildungszwecken den Ablauf von Lisp-Programmen abbildet. Nachträgliches Editieren ist hier möglich, aber nicht notwendig. Eine Theorie zur Auswahl von Animationstechniken zur Visualisierung von Programmabläufen oder -strukturen wird nicht vorgestellt, wenn auch mit der Typisierung von Aktionsprimitiven in BAECKER 1979 ein erster Schritt in diese Richtung getan wurde. Eine Methode, kontinuierliche Bewegungen durch Überlagerung von Bitmaps darzustellen, wird in COOK 1983 entwickelt, um auch auf kleinen Rechnern dynamische Graphik implementieren zu können.

Die hohen Anforderungen, die an ein solches Bildgenerierungssystem gestellt werden, lassen sich mit herkömmlichen graphischen Kommando-sprachen (exemplarisch NEWMAN/SPROULL 1979) kaum bewältigen, sondern erfordern ein objektorientiertes wissensbasiertes System (ZDYBEL et al. 1981), das sowohl konzeptuelle wie auch graphische Information modelliert. Als Repräsentationsverfahren haben sich Frames (ZDYBEL et al. 1981, BROWN/CHANDRASEKARAN 1981) und semantische Netze (FIRSCHER/FISCHLER 1972) wegen ihrer Flexibilität gegenüber Produktionsregelsystemen (STINY 1975) durchgesetzt (GIUSTINY/LEVINE/MALOWANY 1978).

4.2.2.1.3 Graphisch-interaktive Präsentation von Textinhalten

a) Eignung von Graphik zur Präsentation von Textinhalten

zu betrachten (vgl. exemplarisch KINTSCH/VAN DIJK 1978) bisher kaum in. Für die Notwendigkeit graphisch unterstützter Interaktion führt Hayes zwei Hauptgründe an (HAYES 1980):

- Die Existenz nicht verbaler Kommunikationsformen wird beim natürlichsprachlichen Ansatz vernachlässigt.
- Die Mensch-Maschine-Kommunikation als Analogie zur Mensch-Mensch-Kommunikation aufzufassen heißt, die Stärken und Schwächen dieser Geräte zu vernachlässigen.

Eine Mischung von graphischer und verbaler Kommunikation bei wechselndem Abstraktionsgrad erscheint außerdem als geeignet, individuellem Wahrnehmungsverhalten gerecht zu werden, das sich nach ALESANDRINI et al. 1984 in die Kategorien analytisch vs holistisch und verbal vs visuell

einteilen läßt, wobei Information über Textstrukturen besonders gut graphisch vermittelt werden kann (STRONG 1974). Die für das Textverständnis wichtigen Kohäsions- und Kohärenzstrukturen sind die Grundlage für den Aufbau von Textgraphen, die den von RAUCH 1979 geforderten Ansatz des interaktiven Abstracting unterstützen.

b) Kriterien zur kognitiv und ergonomisch günstigen Präsentation von Texten

Die empirischen Untersuchungen, die von der Psychologie zu diesem Themenkreis angestellt wurden, sind nur sehr schwer für konkrete Designentscheidungen nutzbar zu machen. Sie befassen sich fast nur mit den physischen Aspekten der Darstellung, wie z.B. der Geschwindigkeit (BEVAN 1981) des Bildaufbaus. Dieser Eindruck wird von WAERN/ROLLENHAGEN 1983 bestätigt, einem knappen State-of-the-Art Bericht zum Thema Textpräsentation. Die dort vorgestellten Untersuchungsergebnisse, die sich mit inhaltlichen Aspekten des Lesens befassen, sind noch nicht operationalisierbar. Welchen Einfluß das Angebot verschiedener Präsentationsformen auf das Textverständnis haben könnte, wurde bislang nicht untersucht. Die Verfahren, die in TOPOGRAPHIC zur Textpräsentation entwickelt werden, lassen sich daher nicht direkt aus empirisch ermittelten Forschungsergebnissen ableiten. Andererseits erlauben die Ergebnisse von KENNEDY 1983, der aus Augenbewegungen beim Lesen schließt, daß beim Textverstehen auch räumliche Informationen (die Position eines Wortes) gespeichert werden, den Schluß, daß eine strukturierte graphische Darstellung den Verständnisprozess fördern könnte.

c) Kaskadiertes Abstracting, Zooming

In Analogie zur photographischen Optik wurden in der Bildverarbeitung und Computer Animation Zooming-Verfahren entwickelt, graphische Objekte bei verkleinertem Bildausschnitt in größerem Detailierungsgrad zu präsentieren (BAECKER 1979).

Darüber hinaus wurde das Konzept des Zooming zu einer Kombination von Präsentations- und Retrievalwerkzeug entwickelt, das nicht nur richtungs- und ausschnittsorientierte Vergrößerung sondern einen inhaltsbezogenen Zugriff auf Bilddetails erlaubt. In CHANG et al. 1980 werden daraus drei Zoom-Operatoren zum Zugriff auf eine Bilddatenbasis abgeleitet :

- vertical Zoom : Ausschnittsvergrößerung
- horizontal Zoom : Inhaltsgesteuerter Zugriff auf Bilddetails
- diagonal Zoom : bildübergreifender horizontal Zoom

Mit dem "Dynamic Book" (WEYER 1982) werden ähnliche Mechanismen zum Faktenretrieval in Freitexten eingesetzt. Je nach Interesse und Vorwissen des Benutzers kann hier ein Text, in dem Beispiel ein Geschichtsbuch, aus verschiedenen Sichten dargestellt werden, wobei die Textform der jeweiligen Perspektive angepaßt wird. Lehrbuchartige Textpassagen (England im 18. Jahrh.) können sich mit enzyklopädischer Präsentation (englische Könige) abwechseln. Das Problem der Wissensgewinnung aus unterschiedlichen Texten bleibt jedoch unberücksichtigt.

L i t e r a t u r v e r z e i c h n i s

- ALESANDRINI, K. / J. LANGSTAFF / M. WITTROCK (1984)
Visual-Verbal and Analytic-Holistic Strategies, Abilities and Styles.
In: Journal of Education Research 77. 1984, No.3, pp.151-157.
- BAECKER, R. (1975)
Two Systems which Produce Animated Representations of the Execution of
Computer Programs. In: SIGCSE Bull. 7. 1975, No.1, pp.158-167,
zitiert nach DIONNE/MACKWORTH 1978.
- BAECKER, R. (1979)
Digital Video Display Systems and Dynamic Graphics. (SIGGRAPH '79
Proceedings). In: Computer Graphics 13. 1979, No.2, pp.48 - 56.
- BERTIN, J. (1982)
Graphische Darstellungen und die graphische Weiterverarbeitung der
Information. Berlin etc.: de Gruyter, 1982.
- BEVAN, N. (1981)
Is there an Optimum Speed for Presenting Text on a VDU? In: Int. J.
Man-Machine-Studies 14. 1981, pp. 59-76.
- BROWN, D. / B. CHANDRASEKARAN (1981)
Design considerations for picture production in a natural language
graphics system. In: Computer Graphics 15. 1981, No.2, pp.174-207.
- CAHN, D. (1980)
Computer-Aided Visualization of Database Structural Relationships. In:
Proc. of the Annual Meeting of the ASIS, 17. 1980, pp.358-360.
- CHANG, S. / B. LIN / R. WALSER (1980)
A Generalized Zooming Technique for Pictorial Database System. In:
CHANG / FU, Pictorial Information Systems. Berlin etc.: Springer,
1980, pp. 257-269.
- COOK, S. (1983)
Playing Cards on the PERQ: An Algorithm for Overlapping Rectangles.
In: Software-Practice and Experience 13. 1983, pp.1043-1053.
- DIONNE, M. / A. MACKWORTH (1978)
ANTICS: A System for Animating LISP-Programs. In: Computer Graphics
and Image Processing 7. 1978, pp.105-119.
- DEHNING, W. / H. ESSIG / S. MAASS (1981)
The Adaptation of Virtual Man-Computer Interfaces to User Requirements
in Dialogs. Berlin etc.: Springer, 1981. (= Lecture Notes in Computer
Science 110)
- FIRSCHEIN, O. / M. FISCHLER (1972)
A Study in Descriptive Representation of Pictorial Data. In: Pattern
Recognition 4. 1972, pp.361-377.
- GIUSTINI, R. / M. LEVINE / A. MALOWANY (1978)
Picture Generation Using Semantic Nets. In: Computer Graphics and
Image Processing 7. 1978, No.1, pp.1-29.

- GMD (1982)
Modellierung existierender Schnittstellen mit Netzen. St. Augustin:
GMD, 1982 (= GMD-Studien Nr. 69)
- HAYES, P. (1980)
Expanding the Horizons of Natural Language Interfaces. In: 18th Annual
Meeting of the ACL, 1980, pp.71-74.
- JAGODZINSKI, A. (1983)
A Theoretical Basis for the Representation of Online Computer Systems
to Naive Users. In: Int. J. of Man-Machine-Studies 18. 1983,
pp.215-252.
- KENNEDY, A. (1983)
On Looking into Space. In: K. Rayner (ed): Eye Movements in Reading.
New York: Academic Pr., 1983, pp.237-251.
- LEWIS, J. (1975)
Tree: An interactive System for Editing Tree Structures. In: Computers
& Graphics 1. 1975, No.1, pp.65-68.
- MORIK, K. (1984)
Partnermodellierung und Interessenprofile bei Dialogsystemen der
künstlichen Intelligenz. Universität Hamburg, Forschungsstelle f.
Inf.wiss. u. Künstl. Intelligenz, 1984. (= Report ANS-24)
- NEWMAN, W. / R. SPROULL (1979)
Principles of Interactive Computer Graphics. New York: McGraw-Hill,
1979.
- NIEVERGELT, J. / J. WEYDERT (1979)
Sites, Modes and Trails: telling the User of an Interactive System
where he is, what he can do, and how to get to Places. Zürich: ETH,
1979. (= ETH-Bericht Nr. 28)
- NIEVERGELT, J. (1982)
Errors in Dialog Design and How to Avoid Them. In: International
Zurich Seminar, 1982, pp.199-205.
- OBERQUELLE, H. (1981)
Communication by Graphic Net Representation. Universität Hamburg, 1981
(=Bericht Nr. 75)
- POLLITT, A. (1981)
An Expert System as an Online Search Intermediary. In: 5th Int.
Online Information Meeting, 1981, pp.25-32.
- POWELL, M. / M. LINTON (1983)
Visual Abstraction in an Interactive Programming Environment. In: ACM
SIGPLAN Notices 18. 1983, No.6, pp.14-21.
- RAUCH, W. (1979)
Automatic Extracting as an Interactive Process. In: Journal of
Cybernetics 9. 1979, pp.103-112.
- SEGLER, H. (1984)
Programmieren mit graphischen Mitteln: Die Überwachung der Ausführung
von GRADE-Programmen am graphischen Bildschirm. In: U. Amman (ed):
Programmiersprachen und Programmentwicklung. Berlin etc.: Springer,
1984, pp.193-206. (= Informatik-Fachberichte 77)

- STAUFER, M. (1984)
Piktogramme für Bürocomputer. Diplomarbeit, Universität Erlangen,
Psychologisches Institut, 1984.
- STELOVSKY, J. (1984)
The Use of Grammars in an Interactive Operating System. In: Notizen
zum Interaktiven Programmieren 12. 1984, pp.33-41.
- STINY, G. (1975)
Pictorial and Formal Aspects of Shape and Shape Grammars. Basel etc.:
Birkhäuser, 1975.
- STRONG, S. (1974)
An Algorithm for Generating Structural Surrogates of English Text. In:
Journal of the ASIS 1. 1974, pp.10-24.
- WAERN, Y. / C. ROLLENHAGEN (1983)
Reading Text from Visual Display Units (VDUs). In: Int. J.
Man-Machine-Studies 18. 1983, pp.441-465.
- WEYER, S. (1982)
The Design of a Dynamic Book for Information Search. In: Int. J.
Man-Machine-Studies 17. 1982, pp.87-107.
- ZDYBEL, F. / N. GREENFELD / M. YONKE / J. GIBBONS (1981)
An Information Presentation System. In: Proc. of the 7th Int. Joint
Conf. on AI, Vancouver 1981, pp.978-984.
- ZHANG, Q. / A. MENDELZON (1983)
A graphical Query Language for Entity-Relationship. In: C.G. Davis et
al. (ed): Entity-Relationship Approach to Software Engineering.
Amsterdam: North-Holland, 1983, pp.441-448.

4.2.2.2 Alternative Formen des Retrieval

Rein formalsprachliche Zugangssysteme sind für Benutzer von Informationssystemen Hemmnisse, indem sie ihm Kenntnis von abstrakten Sprachkonstrukten und Systemstrukturen abverlangen. Deshalb sind in der letzten Zeit zahlreiche Ansätze entwickelt worden, das konventionelle boolsche Retrieval über Funktionserweiterungen hinaus, wie sie z.B. in TESKEY 1983 vorgetragen werden, qualitativ zu verbessern. Dabei können zwei Hauptansatzpunkte unterschieden werden:

4.2.2.2.1 Front-End-Systeme

Front-End-Systeme stellen eine benutzerfreundlichere Zugangssprache mit einfacherer Syntax zur Verfügung. Die innere Struktur des Systems und damit seine Funktionalität bleibt davon jedoch unbetroffen.

a) natürlichsprachliche Front-Ends

Die Forderung nach natürlichsprachlicher Frageformulierung begründet sich in erster Linie aus der Annahme, daß der Benutzer seine Interessen am leichtesten mit dem Kommunikationsmedium formulieren könne, das ihm von Natur her am vertrautesten ist (HARRIS 1977). Aufbauend auf experimentellen Systemen, wie LUNAR (WOODS/KAPLAN/NASH-WEBBER 1972) und PLANES

(WALTZ 1978), wurden umfangreiche Produktionssysteme erstellt, wie LIFER (HENDRIX et al. 1978) und ROBOT (HARRIS 1977) für den Zugriff auf große Datenbanken oder MASQUERADE (BRZOZOWSKI 1983) und FIRST (DATTOLA 1979) im Bereich des Referenzretrieval. Obwohl diese Systeme den Benutzer von den von der Mühe entlasten, seine Anfrage in einer syntaktisch unvertrauten Sprache zu formulieren, nehmen sie ihm doch nicht die Verantwortung, die Frage in ihrer Begrifflichkeit auf die Struktur der jeweiligen Datenbasis anzupassen. Die Fähigkeit zu Metadialogen ist bei diesen Systemen sehr eingeschränkt, da sie nur wenig explizites Wissen über die Struktur der jeweils zugrunde liegenden Datenbasis zur Verfügung haben.

b) graphisches Retrieval

Extensives Wissen über die Struktur der Daten vermitteln die graphischen Retrievalsprachen, die auf die Vertrautheit des Mediums verzichten, um strukturelle Information als Assoziationshilfe bei der Frageformulierung anbieten zu können. In den ersten Systemen (z.B. CUPID) bestand die Unterstützung allein in der Vorgabe formularähnlicher graphischer Strukturen (McDONALD 1975). Mit der Möglichkeit der interaktiven Navigation in Datenbankschemata, z.B. mit gql/ER (ZHANG/MENDELZON 1983), oder Thesauri, z.B. mit CALIBAN (FREI/JAUSLIN 1982), wurde eine neue Qualität graphischen Retrievals erreicht. Andererseits werden die Ausdrucksmöglichkeiten graphischer Interaktion von CALIBAN nicht ausgeschöpft. Die Einschränkung der Darstellung auf Bäume erschwert eine Navigation, die nicht zielgerichtet "top-down" erfolgt, sondern im Sinne des "Browsing" dem Gewinnen von Übersicht dienen soll. Die Formulierung der Suchanfrage erfolgt auf tabellarischem Wege. Genuin graphische Stilmittel, die vielleicht sogar eine erweiterte Funktionalität erlauben würden, kommen hier nicht zum tragen.

c) Spatial Data Management

Gegen die oben geschilderten Methoden des natürlichsprachigen bzw. graphischen Retrieval grenzen sich die Entwickler des Spatial Data Management Systems SDMS (HEROT et al. 1980) explizit ab. Sie formulieren ein neues Paradigma der Informationsvermittlung, das nicht nur den Gesichtssinn des Menschen, sondern auch das Gehör und den Tastsinn in den Vermittlungsprozeß einschließt (DONELSON 1978). Als Stilmittel werden animierte Bildsequenzen und situationsspezifische Geräusche eingesetzt. Dieser Ansatz ist auf die Problematik des Textretrieval nicht unmittelbar anzuwenden, liefert jedoch einige wichtige Anregungen (wie z.B. den Einsatz von Animation).

4.2.2.2 Zugriffsheuristiken

s

weniger der syntaktische als der semantisch pragmatische Aspekt der Suchanfrage steht im Mittelpunkt eines weiteren Lösungsansatzes, der in erster Linie darauf abzielt, Strategien zu entwickeln, eine Anfrage so zu formulieren, daß sie die Wünsche des Benutzers möglichst genau wiedergibt. Von Heuristiken ausgehend, wie dem "Retrieval by Example" (THOMAS/GOULD 1975) und dem "Retrieval by Reformulation" (TOU et al. 1982), das eine interaktive Verfeinerung der Suche anhand von Zwischenergebnissen ermöglicht, werden Expertensysteme für das Information Retrieval realisiert, die den Benutzer bei der Auswahl von Datenbasen und dem Eingrenzen des Suchraums unterstützen (MARCUS 1981). Das Systemwissen, das dazu benötigt wird, läßt sich nach POLLITT 1981 in folgende Kategorien einteilen:

- Kenntnis der Systemstruktur
- Retrievalstrategien
- Fachwissen aus dem Diskursbereich
- Wissen über den individuellen Benutzer.

4.2.2.2.3 Kombinierte Systeme

Natürlich wurden in der Folge Systeme realisiert, die beide Möglichkeiten zur Verbesserung des Datenzugriffs integrieren. Für die natürlichsprachlichen Zugriffssysteme bedeutet dies, bisher getrennte Wissensbasen, wie Sprachwissen und Wissen über den Diskursbereich, zusammenzuführen und für Inferenzen zu nutzen. Diese Entwicklung führte im Bereich der Datenbanken zu den natürlichsprachlichen Informationssystemen, wie HAM-ANS (NEBEL/MARBURGER 1982) und beim Information Retrieval zum 'Conceptual Retrieval' von SCHANK/KOLODNER/DE JONG 1981.

Daß diese Integration von Interface und Expertensystemkomponente im Bereich des graphischen Retrieval nicht in dieser Konsequenz vollzogen wurde, mag daran liegen, daß Graphikinformation im allgemeinen auf einem niedrigeren konzeptuellen Niveau modelliert wird als sprachliche. In diesem Zusammenhang ist das von BROWN/CHANDRASEKARAN 1981 vorgestellte interaktive Graphiksystem mit natürlichsprachlichem Zugang von großem Interesse, da hier eine Integration beider Wissensbasen vorgenommen wurde.

L i t e r a t u r v e r z e i c h n i s

- BRZOZOWSKI, J. (1983)
MASQUERADE: Searching the Full Text of Abstracts Using Automatic Indexing. In: Journal of Information Science 1983, No.6, pp.67-73.
- BROWN, D. / B. CHANDRASEKARAN (1981)
Design considerations for picture production in a natural language graphics system. In: Computer Graphics 1981, No.2, pp.174-207.
- DATTOLA, R. (1979)
FIRST: Flexible Information Retrieval System for Text. In: Journal of the ASIS, 1979, pp.9-14.
- DONELSON, W. (1978)
Spatial Management of Information. In: SIGGRAPH - Computer Graphics 12. 1978, No.3, pp.203-209.
- FREI, H.P. / J.F. JAUSLIN (1983)
Graphical Presentation of Information and Services: A User-oriented Interface. In: Information Technology: Research and Development 2. 1983, pp.23-42.
- HARRIS, L. (1977)
User oriented data base query with the ROBOT natural language query system. In: Int. J. Man-Machine-Studies 9. 1977, pp 697-713.
- HENDRIX, G. / E. SACERDOTI / D. SAGALOWICZ / J. SLOCUM (1978)
Developing a natural language interface to complex data. In: ACM Transactions on Database Systems, 1978, No.3, pp. 105-147.

- HEROT, C. / R. CARLING / M. FRIEDEL / D. KRAMLICH (1980)
A Prototype Spatial Information System. In: Computer Graphics 14.
1980, No.3, pp.63-70.
- MARCUS, R. (1981)
An Automated Expert Assistant for Information Retrieval. In: American
Society for Information Science 1981, pp. 270-273.
- MARTIN, T. (1980)
Information Retrieval. In: H. SMITH / T. GREEN (eds): Human
Interaction with Computers, London: Academic Pr., 1980, pp.161-175.
- MCDONALD, N. (1975)
CUPID: A Graphics Oriented Facility for Support of Nonprogrammer.
Interactions with a Data Base. University of California at Berkeley
(Ph.D.Diss.), zitiert nach MARTIN 1980.
- NEBEL, B. / H. MARBURGER (1982)
Das natürlichsprachliche System HAM-ANS: Intelligenter Zugriff auf
heterogene Wissens- und Datenbasen. Universität Hamburg, Forschungs-
sstelle f. Inf.wiss. u. Künstl. Intelligenz, 1982. (= Bericht ANS-7)
- POLLITT, A. (1981)
An Expert System as an Online Search Intermediary. In: 5th Int.
Online Information Meeting, 1981, pp.25-32.
- SCHANK, R. / J.KOLODNER / G. DeJONG (1981)
Conceptual Information Retrieval. In: ODDY, R. / S. ROBERTSON / C.
van RIJSBERGEN / P. WILLIAMS (eds): Information Retrieval Research,
London: Butterworths, 1981, pp.94-116.
- TESKEY, F.N.
Implementing the Basic Functions of Free Text Information Retrieval
Using Binary Relationships. In: Journal of Information Science 6.
1983, pp.165-172.
- THOMAS, J. / J. GOULD (1975)
A Psychological Study of Query by Example. In: National Computer
Conference Proceedings, Montvale/NJ: AFIPS Pr., 1975, pp.439-446.
- TOU, F. / M. WILLIAMS / R. FIKES / A. HENDERSON / T. MALONE (1982)
RABBIT: An Intelligent Database Assistant. In: AAAI-82, 1982, pp.
314-318.
- WALTZ, D. (1978)
An English Language Question Answering System for a large relational
Database. In: Communications of the ACM 21. 1978, pp.526-539.
- WOODS, W. / R. KAPLAN / B. NASH-WEBBER (1972)
The LUNAR Sciences Natural Language Information System: Final Report.
Cambridge, Mass.: Bolt, Beranek and Newman, 1972. (= Report No.
2378)
- ZHANG, Q. / A. MENDELZON (1983)
A Graphical Query Language for Entity-Relationship. In: C.G. Davis et
al. (ed): Entity-Relationship Approach to Software Engineering.
Amsterdam etc.: North-Holland, 1983, pp. 441-448.

5. Kurzberichte zu den Vorläuferprojekten

5.1 Kurzbericht zum Projekt TOPIC I

Bei Projektende (02/85) wird mit TOPIC I ein in Grundzügen lauffähiges Experimentalsystem zur Verfügung stehen, bestehend aus:

- (1) einem rudimentären Wortexperten-Parser (ca. 20-50 Wortexperten), dessen Rahmensystem bis auf einige pathologische Randbedingungen (insb. bzgl. Inter-Wortexperten-Kommunikation) voll funktionsfähig sein wird
- (2) einem rudimentären Weltwissenssystem (ca. 150-300 Frames, bei durchschnittlich 3 Slots/Frame), dessen Verwaltungskomponente vollständig realisiert sein wird
- (3) eigens entwickelten infrastrukturellen Software-Werkzeugen, wie Wortexperten-Editor und Weltwissens-Editor
- (4) einem groben morphologischen Reduktionssystem, das z.T. schon über Routinen zur Nominaldekomposition verfügt
- (5) einer Textkondensierungskomponente zum Aufbau von Textgraphen aus Weltwissensdaten sowie einem entsprechenden Textwissensverwaltungssystem
- (6) aufgrund von Adressraumengpässen einer Inter-Prozesskommunikation zwischen System-Monitor, Wortexperten-, Welt- und Textwissenssystem

TOPIC I wird einem Test bei einer Folge von vorläufigen Textanalysen (ca. 10 Texte) unterzogen, um seine Funktionsfähigkeit im Sinne der störungsfreien Kommunikation der oben beschriebenen Komponenten zu demonstrieren.

5.2 Kurzbericht zum Projekt TOPOGRAPHIC I

Bis zum Projektende (02/85) wird ein prototypisches graphisches Interface implementiert, das die Entwicklungsarbeiten am TOPIC-System unterstützt und deswegen auf die Funktion des "Knowledge Engineering" ausgerichtet ist. Im einzelnen werden folgende Projektziele realisiert sein:

1. experimentelle Visualisierung von Weltwissensstrukturen, Textgraphen und Wortexperten
2. Unterstützung der TOPIC-internen Editoren durch:
 - a) graphisch-interaktive Manipulation des Weltwissens
 - b) Erproben von Varianten der Benutzerführung auf der Basis des "General Purpose Interface" XS2 der ETH Zürich
3. Kopplung der Lilith an den TOPIC-Rechner ONYX C8002

Die bei der Implementation des Prototyps gewonnenen Erfahrungen werden zur Verfeinerung der Konzeption von Bildschirmlayout und Dialogführung genutzt. Mit Hilfe der Graphiksoftware der ETH Zürich wurden bereits erste Layoutstudien für das Oberflächendesign des projektierten Systems entworfen (Abb. 1-4).

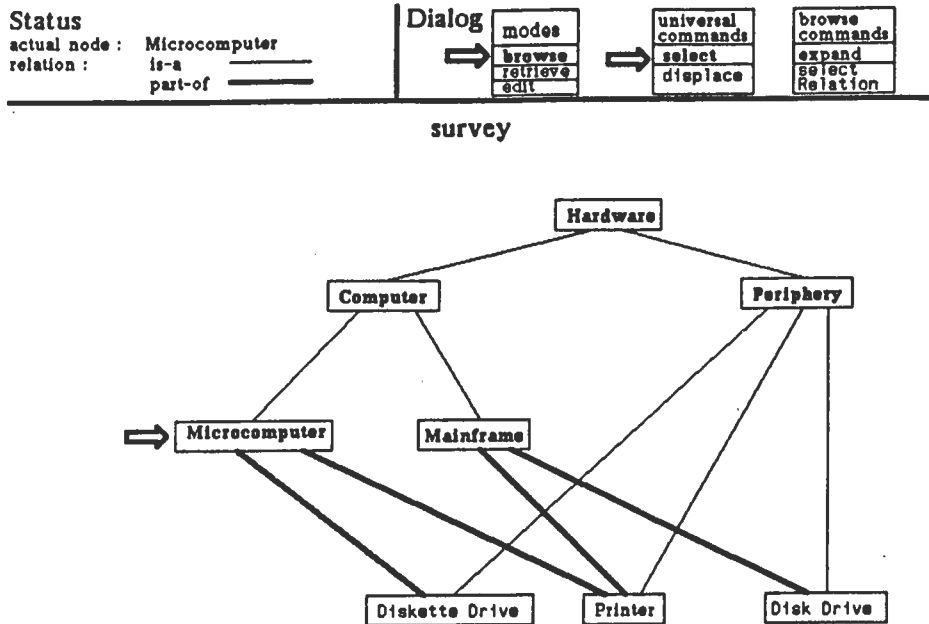


Abb. 1

Eine typische Navigationssituation (vgl. TZ 6) zeigt Abb. 1. Ein Teilnetz aus dem Weltwissen informiert über den konzeptuellen Kontext des aktuellen Knotens ("Microcomputer"), wobei die Oberbegriffe und Komponenten des Begriffs (in verschiedenen Relationsverkettungen) erscheinen. Das angewählte Konzept wird kontrastiert durch eine alternative Spezialisierung ("Mainframe") des Oberbegriffs ("Computer"). Sowohl Gemeinsamkeiten als auch Unterschiede der beiden Konzepte deutet die "part_of"-Relation an. Die Darstellung der Relationen wird in der "Status"-Beschreibung festgelegt, als weitere Information bietet das rechte obere Fenster den Dialogzustand an.

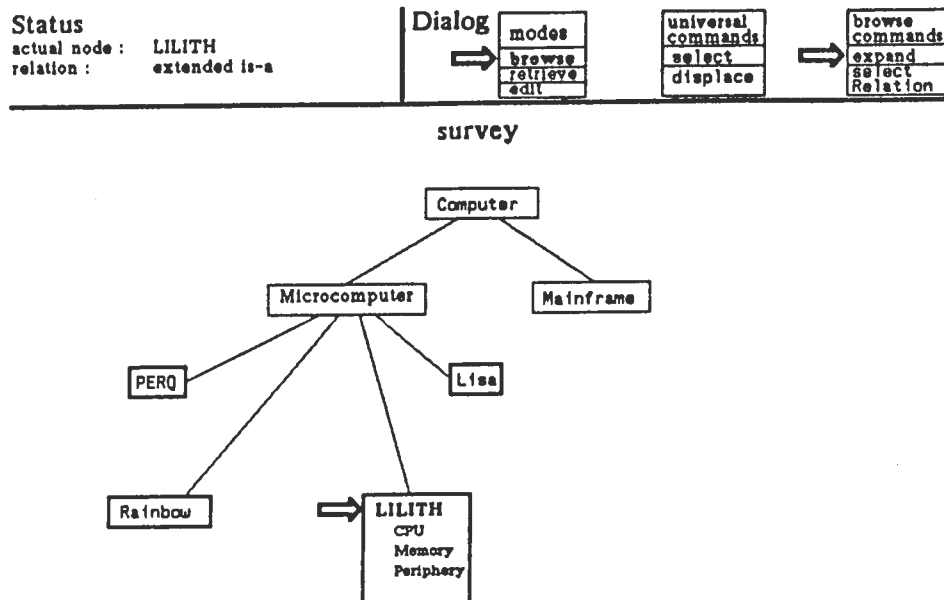


Abb. 2

Durch Expansion eines Frames wird die zugehörige Detailinformation (Slots) sichtbar und liefert Entscheidungshilfen für den Benutzer, der entweder das "Browsing" fortsetzen kann (durch Anwahl anderer Knoten), oder ein "Zooming" des aktuellen Begriffs einleitet (Abb. 3).

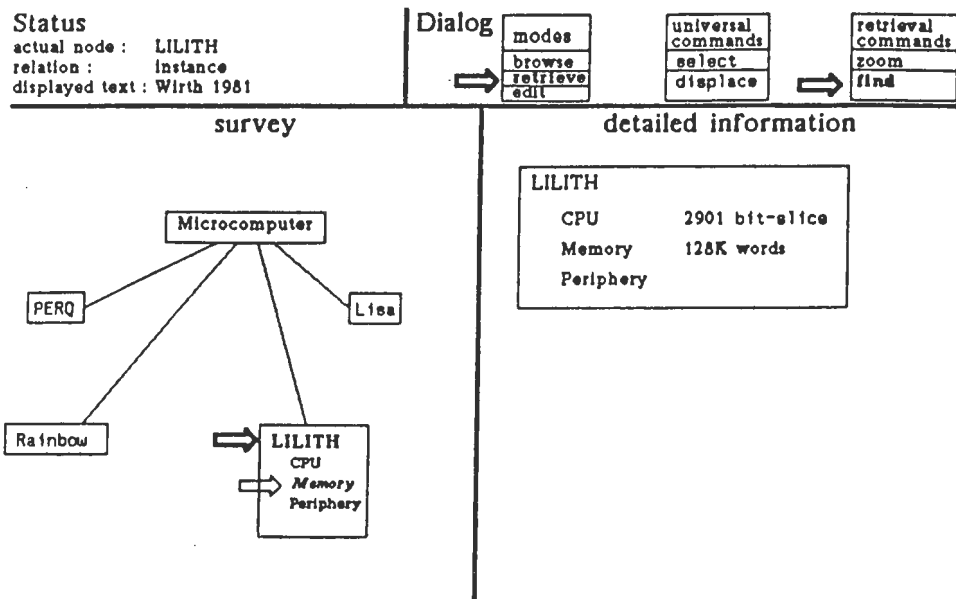


Abb. 3

Das "Zooming" eines Frames führt zunächst zu tabellarischer Fakteninformation. Im Statusfeld erscheint jedoch der Hinweis, daß jetzt bereits Textwissen dargestellt wird.

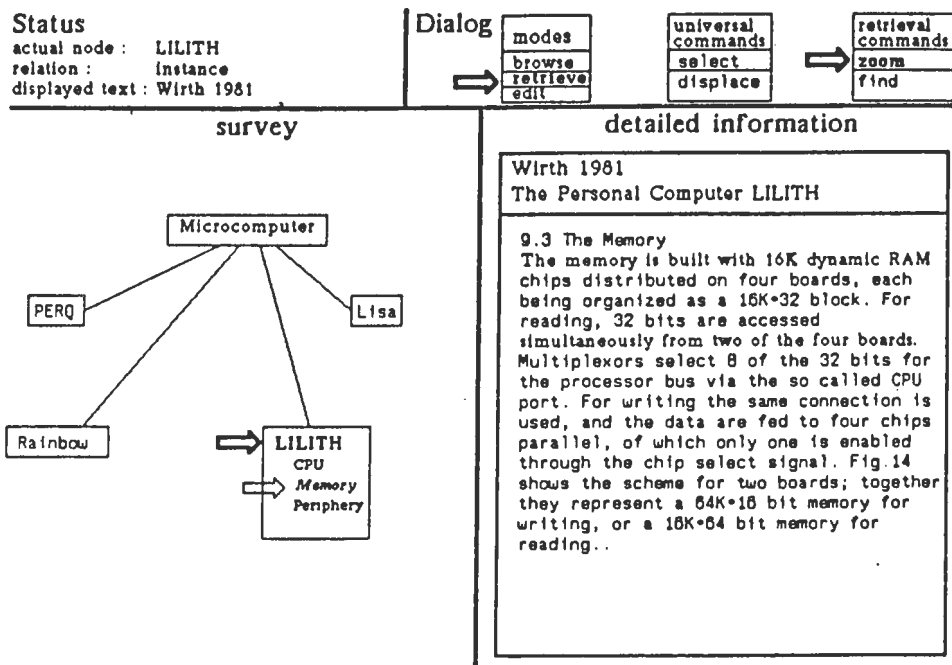


Abb. 4

Als letzte Stufe des "kaskadierten Abstracting" bietet das System die zu einzelnen Textkonstituenten oder Frames gehörigen Textfragmente in variabler Ausführlichkeit an. Abb. 4 zeigt ein solches Stadium des Retrievaldialogs.

5.3 Veröffentlichungen aus den Projekten TOPIC und TOPOGRAPHIC

a) Externe Veröffentlichungen

HAHN, Udo

Textual Expertise in Word Experts: An Approach to Text Parsing Based on Topic/Comment Monitoring.

In: COLING-84: Proceedings of the 10th International Conference on Computational Linguistics & the 22nd Annual Meeting of the Association for Computational Linguistics, Stanford, CA, USA, July 2-6, 1984. ACL, 1984, pp. 402-408.

HAHN, Udo / REIMER, Ulrich

Informationslinguistische Konzepte der Volltextverarbeitung in TOPIC.

In: Deutscher Dokumentartag 1982. Lübeck-Travemünde, 27.9.-30.9.1982. München u.a.: Saur, 1983, S. 345-385.

TOPIC: A System for Automatic Text Condensation.

In: ACM-SIGART Newsletter, No. 83, Jan. 1983, pp. 20-21.

Automatisches Textkondensierungssystem TOPIC.

In: Rundbrief der Fachgruppe Künstliche Intelligenz in der Gesellschaft für Informatik, 1983.

In: LDV Fittings Info, 1983, pp. 33-34.

Wortexperten-Parsing: Text-Parsing mit einer verteilten lexikalischen Grammatik im Rahmen des automatischen Textkondensierungssystems "TOPIC".

In: Linguistische Berichte, Nr. 88, Dec. 1983, S. 56-78.

Heuristic Text Parsing in "TOPIC": Methodological Issues in a Knowledge-Based Text Condensation System.

In: H.J. Dietschmann (ed): Representation and Exchange of Knowledge as a Basis of Information Processes. IRFIS 5: Proceedings of the 5th International Research Forum in Information Science, Heidelberg, FRG, September 5-7, 1983. Amsterdam u.a.: North-Holland, 1984, pp. 143-163.

Computing Text Constituency: An Algorithmic Approach to the Generation of Text Graphs.

In: C.J. van Rijsbergen (ed): Research and Development in Information Retrieval. Proceedings of the 3rd Joint BCS and ACM Symposium. Cambridge, England, 2-6 July 1984. Cambridge u.a.: Cambridge Univ. Pr., 1984, pp. 343-368.

KUHLEN, Rainer

Volltextanalyse zum Zwecke des Abstracting.

In: Deutscher Dokumentartag 1982, Lübeck-Travemünde, 27.9.-30.9.1982. München u.a.: Saur, 1983, S. 300-312.

Some Similarities and Differences between Intellectual and Machine Text Understanding for the Purpose of Abstracting.

In: H.J. Dietschmann (ed): Representation and Exchange of Knowledge as a Basis of Information Processes. IRFIS 5: Proceedings of the 5th International Research Forum in Information Science, Heidelberg, FRG, September 5-7, 1983. Amsterdam u.a.: North-Holland, 1984, pp. 87-109.

REIMER, Ulrich

A System-Controlled Multi-Type Specialization Hierarchy.

In: L. Kerschberg (ed): Proceedings of the First International Workshop on Expert Database Systems, Kiawah Island, South Carolina, USA, Oct 24-27, 1984. Vol 2.

Columbia, SC: Univ. of South Carolina, Inst. of Information Management, Technology and Policy, 1984, pp. 510-521.

REIMER, Ulrich / HAHN, Udo

A Formal Approach to the Semantics of a Frame Data Model.

In: IJCAI-83: Proceedings of the 8th International Joint Conference on Artificial Intelligence. 8-12 August 1983, Karlsruhe, W. Germany. Vol. 1. Los Altos, CA: Kaufmann, 1983, pp. 337-339.

On Formal Semantic Properties of a Frame Data Model.

In: Computers and Artificial Intelligence
(forthcoming)

b) Interne Projektberichte ("Blaue Reihe")

HAHN, Udo / KUHLEN, Rainer / REIMER, Ulrich

Konzeption und Aufbau des automatischen Textkondensierungssystems TOPIC. Universität Konstanz, Informationswissenschaft, September 1982.
(Bericht TOPIC-1/82)

HAHN, Udo / REIMER, Ulrich

Informationslinguistische Konzepte der Volltextverarbeitung in TOPIC: Text-Parsing - Frame-basierte Wissensrepräsentation - Kondensierungstransformationen und grafische Visualisierungstechniken. Universität Konstanz, Informationswissenschaft, November 1982.
(Bericht TOPIC-2/82)

REIMER, Ulrich / HAHN, Udo

A Formal Approach to the Semantics of a Frame Data Model. Extended Version.

Universität Konstanz, Informationswissenschaft, Juli 1983.
(Bericht TOPIC-3/83)

HEESE, Claudia

Die Konzeption einer interaktiven graphischen Mensch-Maschine-Schnittstelle für das System TOPIC: TOPOGRAPHIC.

Universität Konstanz, Informationswissenschaft, Projekt TOPOGRAPHIC, August 1983./83)
(Bericht TOPIC-4/83)

HAHN, Udo / REIMER, Ulrich

Heuristic Text Parsing in TOPIC: Methodological Issues in a Knowledge-Based Text Condensation System.

Universität Konstanz, Informationswissenschaft, Oktober 1983, 2nd ed.
(Bericht TOPIC-5/83)

HAHN, Udo / REIMER, Ulrich

Word Expert Parsing: An Approach to Text Parsing with a Distributed Lexical Grammar.

Universität Konstanz, Informationswissenschaft, November 1983.

(Bericht TOPIC-6/83)

KUHLEN, Rainer

Some Similarities and Differences between Intellectual and Machine Text Understanding for the Purpose of Abstracting.

Universität Konstanz, Informationswissenschaft, November 1983.

(Bericht TOPIC-7/83)

HAHN, Udo / REIMER, Ulrich

Computing Text Constituency: An Algorithmic Approach to the Generation of Text Graphs.

Universität Konstanz, Informationswissenschaft, April 1984.

(Bericht TOPIC-8/84)

KUHLEN, Rainer

A Knowledge-Based Text Analysis System for the Graphically Supported Production of Cascaded Text Condensates.

Universität Konstanz, Informationswissenschaft, Mai 1984.

(Bericht TOPIC-9/84)

HAHN, Udo

Textual Expertise in Word Experts: An Approach to Text Parsing Based on Topic/Comment Monitoring. Extended Version.

Universität Konstanz, Informationswissenschaft, Juni 1984.

(Bericht TOPIC-10/84)

REIMER, Ulrich

A System-Controlled Multi-Type Specialization Hierarchy.

Universität Konstanz, Informationswissenschaft, September 1984.

(Bericht TOPIC-11/84)

HAMMWÖHNER, Rainer / THIEL, Ulrich

Graphische Kommunikations- und Präsentationsformen für komplexe Wissens- und Textstrukturen: zur Konzeption eines graphischen Interface für ein wissensbasiertes Textkondensierungssystem.

Universität Konstanz, Informationswissenschaft, August 1984.

(Bericht TOPOGRAPHIC-1/84)

HAMMWÖHNER, Rainer / THIEL, Ulrich

TOPOGRAPHIC: eine graphisch interaktive Retrievalschnittstelle.

Universität Konstanz, Informationswissenschaft, November 1984.

(Bericht TOPOGRAPHIC-2/84)

PROJEKTBERICHTE

- Hahn, U. / Kuhlen, R. / Reimer, U.
Konzeption und Aufbau des automatischen Textkondensierungssystems TOPIC.
Universität Konstanz, Informationswissenschaft, September 1982.
(Bericht TOPIC-1/82)
- Hahn, U. / Reimer, U.
Informationslinguistische Konzepte der Volltextverarbeitung in TOPIC:
Text-Parsing - Frame-basierte Wissensrepräsentation - Kondensierungs-
transformationen und grafische Visualisierungstechniken.
Universität Konstanz, Informationswissenschaft, November 1982.
(Bericht TOPIC-2/82)
- Reimer, U. / Hahn, U.
A Formal Approach to the Semantics of a Frame Data Model. Extended Version.
Universität Konstanz, Informationswissenschaft, Juli 1983.
(Bericht TOPIC-3/83)
- Heese, C.
Die Konzeption einer interaktiven graphischen Mensch-Maschine-
Schnittstelle für das System TOPIC: TOPOGRAPHIC.
Universität Konstanz, Informationswissenschaft, Projekt TOPOGRAPHIC,
August 1983.
(Bericht TOPIC-4/83)
- Hahn, U. / Reimer, U.
Heuristic Text Parsing in TOPIC: Methodological Tools for Knowledge-Based
Text Condensation Systems.
Universität Konstanz, Informationswissenschaft, August 1983.
(Bericht TOPIC-5/83)
- Hahn, U. / Reimer, U.
Word Expert Parsing: An Approach to Text Parsing with a Distributed
Lexical Grammar.
Universität Konstanz, Informationswissenschaft, November 1983.
(Bericht TOPIC-6/83)
- Kuhlen, R.
Some Similarities and Differences between Intellectual and Machine Text
Understanding for the Purpose of Abstracting.
Universität Konstanz, Informationswissenschaft, November 1983.
(Bericht TOPIC-7/83)
- Hahn, U. / Reimer, U.
Computing Text Constituency: An Algorithmic Approach to the Generation
of Text Graphs.
Universität Konstanz, Informationswissenschaft, April 1984.
(Bericht TOPIC-8/84)
- Kuhlen, R.
A Knowledge-Based Text Analysis System for the Graphically Supported
Production of Cascaded Text Condensates.
Universität Konstanz, Informationswissenschaft, Mai 1984.
(Bericht TOPIC-9/84)
- Hahn, U.
Textual Expertise in Word Experts: An Approach to Text Parsing Based on
Topic/Comment Monitoring. Extended Version.
Universität Konstanz, Informationswissenschaft, Mai 1984.
(Bericht TOPIC-10/84)
- Reimer, U.
A System-Controlled Multi-Type Specialization Hierarchy.
Universität Konstanz, Informationswissenschaft, September 1984.
(Bericht TOPIC-11/84)
- Hammwöhner, R. / Thiel, U.
Graphische Kommunikations- und Präsentationsformen für komplexe Wissens-
und Textstrukturen: Zur Konzeption eines graphischen Interface für ein
wissensbasiertes Textkondensierungssystem.
Universität Konstanz, Informationswissenschaft, August 1984.
(Bericht TOPOGRAPHIC-1/84)
- Hammwöhner, R. / Thiel, U.
TOPOGRAPHIC: eine graphisch interaktive Retrievalschnittstelle.
Universität Konstanz, Informationswissenschaft, November 1984.
(Bericht TOPOGRAPHIC-2/84)
- Kuhlen, R.
Informationsvermittlung und Information Management. Die Konzeption des
Aufbaustudiums der Informationswissenschaft an der Universität Konstanz.
Universität Konstanz, Informationswissenschaft, Mai 1982.
(Bericht 1/82)
- Hahn, U.
Leistungsfähigkeit und Grenzen externer publizistischer Datenbanken und
deren Verbesserung durch informationslinguistische Software.
Universität Konstanz, Informationswissenschaft, Juli 1982.
(Bericht 2/82)
- Kuhlen, R.
Das Informationsverhalten an der Universität Konstanz bei der Nutzung
von Datenbanken. Ergebnisse einer gesamtuniversitären Befragung.
Universität Konstanz, Informationswissenschaft, März 1983.
(Bericht 1/83)

- H ü t t e r , B.
Ein Kostenvergleich von Private File Service-Anbietern.
Universität Konstanz, Informationswissenschaft, April 1983.
(Bericht 2/83)
- S t a u d , J. L.
Extrahieren und Verarbeiten numerischer Fakten in Retrieval-Systemen.
Am Beispiel AREMOS.
Universität Konstanz, Informationswissenschaft, Mai 1983.
(Bericht 3/83)
- J a n o v s k y , J. / M ü l l e r - H e i d e n , B.
Organisationstheoretische Konzepte zur Informationsverarbeitung in
Organisationen. Zur theoretischen und empirischen Begründbarkeit einer
informationellen Organisationstheorie.
Universität Konstanz, Informationswissenschaft, November 1983.
(Bericht 4/83)
- K u h l e n , R.
Informationserarbeitung in Organisationen. Zur Rekonstruktion der
Notwendigkeit eines Informationsmanagements in öffentlichen Verwaltungen
und privaten Unternehmungen.
Universität Konstanz, Informationswissenschaft, November 1983.
(Bericht 5/83)
- V o g e l , E.
Zur Konzeption, Realisierung und Akzeptanz des Konstanzer Diplom-
Aufbaustudiengangs Informationswissenschaft.
Universität Konstanz, Informationswissenschaft, November 1983.
(Bericht 6/83)
- R a n k , W.
Ansätze zu einer informationellen Organisationstheorie. Aspekte einer
informierten Organisation.
Universität Konstanz, Informationswissenschaft, November 1983.
(Bericht 7/83)
- S t a u d , J. L.
Integrierte Abspeicherung und Verarbeitung numerischer Daten. Zur
Konzeption eines neuen systemtyps.
Universität Konstanz, Informationswissenschaft, Dezember 1983.
(Bericht 8/83)
- K u h l e n , R.
Kommunikationstechnologien und Organisationsstruktur.
Universität Konstanz, Informationswissenschaft, Dezember 1983.
(Bericht 9/83)
- S t a u d , J. L.
Statistische Verarbeitung von Informationen in Faktenretrievalsystemen.
Am Beispiel der Zeitreihenstatistik in AREMOS.
Universität Konstanz, Informationswissenschaft, März 1984.
(Bericht 1/84)
- V o g e l , E.
Didaktische Hinweise zur Veranstaltungsplanung und -bewertung im Rahmen
des Diplom-Aufbaustudiums Informationswissenschaft.
Universität Konstanz, Informationswissenschaft, März 1984.
(Bericht 2/84)
- K u h l e n , R.
An International Delphi Poll on Future Trends in "Information Linguistic".
Universität Konstanz, Informationswissenschaft, Mai 1984.
(Bericht 3/84)
- S t a u d , J. L.
Computergestützte Informationssysteme in Organisationen: Stand und
Determinanten. Ergebnisse einer empirischen Erhebung im Sommer 1983.
Universität Konstanz, Informationswissenschaft, Oktober 1984.
(Bericht 5/84)
- V o g e l , E.
The Konstanz Conception for a Training Course in Information Science:
Information Counselling and Information Management.
Universität Konstanz, Informationswissenschaft, July 1984.
(Bericht 7/84)
- S t a u d , J. L.
Online Retrieval in Numerical Databases. Peculiarities and Examples.
Universität Konstanz, Informationswissenschaft, September 1984.
(Bericht 8/84)
- H a h n , U.
Expertensysteme als intelligente Informationssysteme: Konzepte für die
funktionale Erweiterung des Information Retrieval.
Universität Konstanz, Informationswissenschaft, November 1984.
(Bericht 9/84)

B e r i c h t zum Projekt Informationsvermittlung für den Zeitraum 1.1.1982
bis 31.3.1983.
Universität Konstanz, Informationswissenschaft, April 1983.

B e r i c h t zum Projekt Informationsvermittlung für den Zeitraum 1.4.83 bis
31.12.83
Universität Konstanz, Informationswissenschaft, Januar 1984.

K o n s t a n z e r K o l l o q u i u m , 1. - 2. Juli 1983.
Bericht zur wissenschaftspolitischen Orientierung der Informations-
wissenschaft.
Universität Konstanz, Informationswissenschaft, Oktober 1983.

K u h l e n , R.
Verfahren und Systeme der Automatischen Übersetzung.
Bericht zum "Tutorial on Machine Translation", 2. - 6. April 1984 in
Lugano, organisiert von ISSCO, Genf.
Universität Konstanz, Informationswissenschaft, April 1984.