

IZ-Arbeitsbericht Nr. 14

**Elektronische Version des Thesaurus und der
Klassifikation Sozialwissenschaften**

Udo Riege

April 1998



InformationsZentrum
Sozialwissenschaften

Lennéstraße 30
D-53113 Bonn
Tel.: 0228/2281-0
Fax.: 0228/2281-120
email: riege@bonn.iz-soz.de
Internet: <http://www.social-science-geis.de>

ISSN: 1431-6943
Herausgeber: Informationszentrum Sozialwissenschaften der Arbeits-
gemeinschaft Sozialwissenschaftlicher Institute e.V.
(ASI)
Druck u. Vertrieb: Informationszentrum Sozialwissenschaften, Bonn
Printed in Germany

Das IZ ist Mitglied der Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. (GESIS), einer
Einrichtung der Wissenschaftsgemeinschaft Gottfried Wilhelm Leibniz (WGL)

Inhalt

1 Einleitung	4
2 Entwicklung des Prototypen	9
2.1 Einbettung in das Projekt GESINE	9
2.2 Grundkonzept für das Oberflächendesign	11
3 Rapid Prototyping und Anwendertests	14
3.1 Entwicklung des Prototypen	14
3.2 Anwendertest Mai 1997	18
3.3 Weiterentwicklung bis Dezember 1997	23
3.3.1 Änderungswünsche, Verbesserungen und Korrekturen	23
3.3.2 Klassifikationsbrowser	24
3.3.3 Umstellung auf Windows 95	25
3.3.4 Sammelbeutel	25
3.3.5 Mehrsprachigkeit	27
3.4 Anwendertest Februar 1998	28
4 Beschreibung Stand März 1998	30
4.1 Starten und Beenden des Programms	31
4.2 Browser-Hauptfenster	31
4.3 Thesaurus	31
4.3.1 Sprachauswahl	32
4.3.2 Eingabefeld für die Suchbegriffe	32
4.3.3 Begriffsliste: Tabpage „Alphabetisch“	33
4.3.4 Begriffsliste: Tabpage „Systematisch“	34
4.3.5 Anzeigefeld	35
4.3.6 Nachweisanzeige	36
4.3.7 Blätter-Buttons	36
4.3.8 Button zur Übernahme	36
4.4 Sammelbeutel	37
4.4.1 Listenfeld für Deskriptoren	38
4.4.2 Fenster nach Übernahme leeren	38
4.4.3 Button zur Übernahme in die Terminalemulation	38
4.4.4 Löschen-Button	38
4.5 Klassifikation	39
4.6 Fenstergröße und -position	41
5 Fazit und Ausblick	41
6 Literatur	43

1 Einleitung

Das InformationsZentrum Sozialwissenschaften (IZ) als Mitglied des Infrastrukturverbundes GESIS* bietet für Wissenschaft, Politik und Gesellschaft verschiedenste Informationsdienstleistungen an. Zu diesem Zweck werden am IZ Datenbanken unter anderem zu Fachliteratur (SOLIS), Forschungsvorhaben und -ergebnissen (FORIS), sowie zu Institutionen aufgebaut und gepflegt. Auf der Basis dieser Datenbanken entstehen zahlreiche Produkte wie die CD-ROM WISO III, gedruckte Informationsdienste, Bibliographien und thematische Dokumentationen. Weiterhin werden die Datenbanken bei kommerziellen DB-Hosts aufgelegt und sind somit online recherchierbar. Recherchen können ebenfalls direkt beim IZ in Auftrag gegeben werden (cf. Krause/Zimmer 1997 als Überblick).

Der Produktionsablauf bei der Herstellung der Datenbanken SOLIS und FORIS beinhaltet eine intellektuelle Erschließung der Dokumente durch Vergabe von Deskriptoren (Schlagworten) des Thesaurus Sozialwissenschaften und Klassifikationsnummern der Klassifikation Sozialwissenschaften. Diese relativ aufwendige Erschließung ist ein wesentliches Qualitätsmerkmal beider Datenbanken.

Bis zum jetzigen Zeitpunkt stehen keine für die Domäne der deutschsprachigen Sozialwissenschaften geprüften Verfahren zur automatischen Inhaltserschließung zur Verfügung. Die derzeitige Diskussion um die Zukunftsperspektiven intellektueller Inhaltserschließung und um die Leistungsfähigkeit maschineller Indexierungs- und Retrievalsysteme stützt sich überwiegend auf im englischsprachigen Raum durchgeführte Studien, allen voran die TREC-Initiative (Text Retrieval Conference) des National Institute for Standards and Technology (NIST) in Gaithersburg (MD, USA). Das derzeit laufende

* GESIS ist die Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V. mit den Mitgliedsinstituten Informationszentrum Sozialwissenschaften (IZ) in Bonn, Zentralarchiv für empirische Sozialforschung an der Universität zu Köln (ZA) und dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) in Mannheim. Die GESIS hat eine Außenstelle in Berlin, deren Schwerpunktaufgabe der Informationstransfer von und nach Osteuropa ist.

GIRT-Projekt des IZ soll den Rahmen für einen aussagefähigen Vergleich moderner, intelligenter Indexierungs- und Retrievalsysteme schaffen (vergl. zum aktuellen Stand der Diskussion und zum GIRT-Projekt Frisch/Kluck 1997).

Aufgrund dieser Tatsachen lag die Aufgabe nahe, das am IZ praktizierte Verfahren der intellektuellen Inhaltserschließung durch geeignete Anwendungssoftware zu unterstützen. Ein elektronisches Werkzeug, welches Thesaurus und Klassifikation Sozialwissenschaft in ihrer gesamten Struktur abbildet und eine Export-Möglichkeit in andere Programme besitzt, wäre zudem sinnvoll auch in anderen Zusammenhängen, bspw. als Kopplung mit einer Recherchesoftware, einsetzbar.

Thesaurus Sozialwissenschaften

Der Thesaurus Sozialwissenschaften ist ein Hilfsmittel für die Arbeit in den Datenbanken SOLIS und FORIS. Bei seiner Entwicklung stand der konkrete Verwendungszweck im Vordergrund: Schaffung eines integrierten Instruments für die inhaltliche Erschließung sozialwissenschaftlicher Literatur und Forschungsarbeiten. Unter Beachtung der DIN-Standards zur Thesauruserstellung war zunächst der Bezug zu SOLIS und FORIS Orientierungsrahmen für alle mit der Entwicklung verbundenen Festlegungen und Entscheidungen, z.B. zum fachlichen Scope oder zum Spezialisierungsgrad der Schlagwörter. Fachlich deckt der Thesaurus derzeit alle Disziplinen der Sozialwissenschaften ab. In gewissem Umfang enthält er auch Begriffe aus angrenzenden Bereichen, um sozialwissenschaftliche Arbeiten mit interdisziplinären oder anwendungsorientierten Fragestellungen angemessen inhaltlich beschreiben zu können.

Der Thesaurus wurde parallel mit den Datenbanken SOLIS und FORIS entwickelt. Ausgangspunkt war eine Liste von ca. 14.000 Deskriptoren, die bei der freien Indexierung von Forschungsarbeiten für die Datenbank FORIS über mehrere Jahre gesammelt wurden und bis 1977 Anwendung fanden. Durch das Zusammenführen von Schreibvarianten und das Streichen von Synonymen und zu speziellen Begriffen konnte die Liste auf ca. 11.000 Deskriptoren reduziert werden. In einem zweiten Entwicklungsschritt wurde 1979 für den Bereich Soziologie dieser Liste ein kontrolliertes und strukturiertes Vokabular mit ca. 4800 Einträgen aufgebaut. Das war ein erster Schritt in Richtung Thesaurus. Im nächsten Schritt entstand 1984 die „Schlagwortliste für die Sozialwissenschaften“, in der die Ergebnisse und

Erfahrungen der vorangegangenen Arbeit einfließen. Diese Schlagwortliste deckte von ihrem fachlichen Umfang her das Gebiet der Sozialwissenschaften insgesamt ab. 1985 erschien dann erstmals der systematische Teil der Schlagwortliste. Für die ergänzten Neuauflagen 1987 und 1993 sind die Begriffsrelationen weiter ausgearbeitet worden, so daß 1996 die Schlagwortliste in den Rang eines „Thesaurus Sozialwissenschaften“ erhoben werden konnte.

Um Nutzern mit geringen Kenntnissen der deutschen Sprache den Zugang zu den Datenbanken des IZ zu erleichtern, wurde die Schlagwortliste bereits 1987 in einer englischen Übersetzung herausgegeben.

In mehrjähriger gemeinsamer Arbeit von IZ und dem Institut für wissenschaftliche Information in den Gesellschaftswissenschaften (INION) der Russischen Akademie der Wissenschaften (Moskau) wurde der Thesaurus Sozialwissenschaften auch ins Russische übersetzt. Ziel war, die sprachlichen Barrieren bei der Suche in deutschen und russischen Datenbanken zu überwinden und den wissenschaftlichen Informationsaustausch zwischen Ost und West zu fördern.

Der deutschsprachige Thesaurus Sozialwissenschaften steht für externe Nutzer bislang nur als zweibändiges Druckwerk zur Verfügung. Band 1 beinhaltet in alphabetischer Auflistung alle zugelassenen 6960 Deskriptoren sowie 3530 Verweise von Nichtdeskriptoren auf Deskriptoren (Stand 1997), Band 2 enthält die systematische Einordnung der Deskriptoren und Nichtdeskriptoren in die 211 Notationen. Erstmals 1987 erschien der deutsch-englische sowie der englisch-deutsche Thesaurus als zweibändiges Druckwerk. 1997 folgte die dreibändige russisch-englisch-deutsche Übersetzung.

Das Vokabular des Thesaurus Sozialwissenschaften wird derzeit über das IZ hinaus von weiteren Informationseinrichtungen und verschiedenen Bibliotheken eingesetzt und hat im Laufe der Jahre zunehmend Leitbildfunktion übernommen.

Klassifikation Sozialwissenschaften

Beim Retrieval in den Datenbanken (inhouse, online oder auf CD-ROM) dienen Klassifikationen vorwiegend der Einschränkung großer Dokumentmengen auf eine disziplinäre Grundrichtung oder einen speziellen Blickwinkel. Aus dieser Funktion läßt sich ableiten, daß eine allzu tiefgehende Differenzierung der Klassen im allgemeinen nicht erforderlich ist. Hinsichtlich des Datenbankinputs am IZ Sozialwissenschaften kam der Nutzung von Klassifi-

kationen keine besondere Bedeutung zu, sie dienten lediglich einer groben, deskriptiven Scopekontrolle und einigen globalen Statistiken. Ferner wurden die Klassifikationen für bibliometrische Analysen genutzt.

Bis 1996 sind unterschiedliche Klassifikationssystematiken für die Datenbanken FORIS und SOLIS angewendet worden. Die Anwendung zweier verschiedener Systematiken mit jeweils wiederum zwei Klassifikationslisten ist durch die ursprünglich getrennt verlaufende Entwicklung von FORIS und SOLIS bedingt gewesen. Die übrigen Erschließungsinstrumente und Erfassungsregeln für beide Datenbanken wurden bereits im Zuge der Umstellung auf das Inhouse-Datenbanksystem aDIS* im Jahre 1987/88 vereinheitlicht bzw. angeglichen. In beiden Klassifikationslisten sind in der Vergangenheit kleinere Überarbeitungen durchgeführt worden. Teilweise sind auch dezierte Untermengen der Datenbestände retrospektiv mit bestimmten Klassifikationsmerkmalen versehen worden.

Im Zuge einer weiteren Vereinheitlichung der Erschließungsinstrumente wurde 1995 eine für die Datenbanken SOLIS und FORIS einheitliche Klassifikation erarbeitet und ab 1996 eingesetzt. Die in vergleichbaren Datenbanken (Sociofile, Psycinfo) angewendeten Klassifikationssysteme, die Listen der Arbeitsgruppen der wissenschaftlichen Gesellschaften (DGS, ISA, DVPW) und einschlägige Literatur wurden mit berücksichtigt.

Insgesamt wurden in der neuen Systematik 158 Klassen geschaffen statt wie bisher 278 für FORIS und 93 für SOLIS. Die Klassen erhielten eine numerische Notation, in der die Gliederung der Hauptklassen und die darunter liegende Hierarchie von Klassen eindeutig abgebildet sind. Die Zahlenschlüssel der Notation sind so angelegt, daß die Klassifikation erweiterungsfähig ist. Die Erfahrungen mit der Anwendung sind sehr positiv. Die vorhandenen Datenbestände von FORIS und SOLIS wurden retrospektiv auf die neue Klassifikation umgesetzt, so daß für den gesamten Datenbestand ein einheitliches Suchinstrument zur Verfügung steht.

* Am IZ wird seit 1989 das Datenbanksystem aDIS (steht für adaptierbares Dokumentations- und Informationssystem) der Münchner Firma aStec verwendet. Als Terminalemulation wurde zunächst ein Produkt der Firma Siemens, später dann EMF®97-WIN von der CONWARE Computer Consulting GmbH eingesetzt.

Anforderungen an die Entwicklung der elektronischen Versionen

Ausgehend von den Erfahrungen bei der Arbeit mit den gedruckten Versionen von Thesaurus und Klassifikation sind in vorbereitenden Gesprächen mit hausinternen Nutzern folgende allgemeine Anforderungen an die Entwicklung einer elektronischen Version formuliert worden:

- Die elektronische Version des Thesaurus Sozialwissenschaften sollte als Nachschlagwerk die alphabetische und systematische Struktur in deutscher und englischer Sprache mit möglichst einfachen Navigationsmöglichkeiten in den Begriffs- und systematischen Hierarchien abbilden. Die Darstellungen sollten bezüglich alphabetischer Liste, Systematik und Sprache in jeder Beziehung kontextsensitiv durchlässig sein, d.h. Benutzereingaben in einem Fenster lösen selbständig Anpassungsvorgänge innerhalb des Fensters bzw. in anderen Fenstern aus (cf. Krause 1995, S. 19 ff.). Das bei umfangreichen bzw. tief gestaffelten Hierarchien oder einem streng hierarchisch aufgebauten Thesaurus oft auftretende Navigationsproblem sollte bei der Gestaltung so gelöst werden, daß auch ein ungeübter Benutzer die Übersicht behält.
- Die elektronische Version der Klassifikation Sozialwissenschaften sollte als Nachschlagwerk die systematische Struktur abbilden. Verweise von Fachbegriffen auf dafür zu verwendende Klassifikationen der bisherigen Index-Liste sollten mit erfaßt und erweitert werden.
- Der Export bzw. die Übernahme von Deskriptoren und Klassifikationen in andere Programme (zunächst in die Recherchekomponente von GESINE, später möglicherweise auch in das am IZ verwendete Dokumentations- und Informationssystem aDIS) sollte möglich sein. Sowohl Thesauruseinträge als auch Klassifikationen eignen sich als ein möglicher Einstieg in eine Datenbankrecherche. Positiv bei dieser Art von Einstieg ist vor allem der starke Vorlagecharakter. Die begriffliche oder hierarchische Gliederung der Datenbestände wird dabei präsentiert und muß somit nicht unbedingt schon bekannt sein. Durch diesen Vorlagecharakter bieten sich diese Zugänge in erster Linie für noch ungeübte Benutzer bzw. für Recherchen in noch unbekanntem Themenbereichen an.

Mit diesen Anforderungen als Ausgangsbasis wurde der Browser für den deutschen und englischen Thesaurus und die Klassifikation Sozialwissen-

schaften entwickelt. Er liegt als eigenständiger Teil des Programmpakets GESINE am Informationszentrum Sozialwissenschaften in Bonn vor.

Speziell für die IZ-interne Nutzung wurde die allgemeine Exportfunktionalität erweitert, um eine komfortable Vergabe und Übernahme von Deskriptoren in aDIS (via Terminalemulation EMF[®]97-WIN) bei der inhaltlichen Erschließung von SOLIS- und FORIS-Dokumenten zu unterstützen.

Im folgenden werden die vorbereitenden Entwicklungsarbeiten (rapid prototyping), die Anwendertests im IZ mit deren Impulsen für die Programmverbesserung sowie die Software in ihrem derzeitigen Stand beschrieben. Unter Einbeziehung und in ständiger Abstimmung mit den fachlich Verantwortlichen für Thesaurus und Klassifikation sowie den potentiellen Anwendern konnte so ein Stand der Entwicklung erreicht werden, der einen Praxiseinsatz zuläßt.

2 Entwicklung des Prototypen

2.1 Einbettung in das Projekt GESINE

Der sozialwissenschaftlich interessierte Nutzer erwartet von der GESIS in zunehmendem Maße die Vermittlung qualitativ hochwertiger und komplexer Informationen über sozialwissenschaftliche Fragestellungen und Ergebnisse sowie über die Struktur und Entwicklung sozialwissenschaftlicher Forschungsfelder. Insbesondere die Verdichtung der ständig wachsenden Informationsmengen wird an Bedeutung gewinnen.

Das Grundproblem ist, daß mit einer konventionellen Datenbankrecherche nur ein Teil dieses „erhöhten“ und qualitativ veränderten Informationsbedarfs abgedeckt werden kann.

Im Rahmen des Projekts GESINE (steht für: die Datenbestände der GESIS integrierendes sozialwissenschaftliches Informationssystem) wird daher ein Informationssystem für die Sozialwissenschaften auf der Basis der heterogenen Datenbestände der GESIS neu konzipiert. Der Benutzer soll durch ein host-übergreifendes Informationsmanagement die Vorstellung vermittelt bekommen, nur mit einer Datenbank zu kommunizieren. Von besonderem Interesse ist hier v.a. die Behandlung strukturierter und unstrukturierter Infor-

mation (Text-Daten-Integration) und die Entwicklung eines domänenadäquaten Recherchemodells.

Für die Datenbestände des Informationszentrums Sozialwissenschaften (Literaturdatenbank SOLIS und Forschungsprojektdatenbank FORIS) sowie für das GESIS-Clearinghouse und die ALLBUS-Bibliographie ist bereits ein GESIS-weites relationales Datenmodell auf der Basis einer ORACLE-Datenbank definiert und implementiert worden. Eine Ausweitung der Datengrundlage hin zur Integration von Fakten-Informationsbeständen (Fragebögen und Daten empirischer Erhebungen) des Zentrums für Umfragen, Methoden und Analysen (ZUMA, Mannheim) und des Zentralarchivs für empirische Sozialforschung (ZA, Köln) ist für die Zukunft geplant.

Im Projekt GESINE wird erstmalig versucht, die im WOB-Modell (Krause 1995) formulierten Gestaltungsrichtlinien für grafische Oberflächen auf den Gegenstandsbereich des Textretrieval anzuwenden bzw. das Modell für diesen Bereich zu ergänzen. Ein allgemeiner Oberflächenentwurf für das Textretrieval kann nur dann stabil sein und später durch Ausweitung auf verschiedene Domänen ohne größere Designbrüche angewendet werden, wenn möglichst viele der später auftretenden Probleme und Suchmöglichkeiten antizipiert und behandelt werden. Aus diesem Grund wird bei der Softwareentwicklung bereits von einem Datenumfang ausgegangen, der in GESINE zunächst noch nicht vollständig vorhanden ist. Neben den Datenbanken SOLIS und FORIS werden im Entwurf daher auch weitere textuelle Informationsbestände wie das Angebot im Internet, die veröffentlichten Arbeitsberichte oder die Jahresberichte berücksichtigt.

Eine Software zur Durchführung von Suchen über Textmengen umfaßt neben der reinen Recherchefunktionalität weitere Bereiche, wie z.B. die Ablage von Suchanfragen und Ergebnissen oder den Export von Daten in Fremdformate. Diese Funktionalitätsbereiche sind zum großen Teil bereits in anderen, auf dem WOB-Modell basierenden Anwendungen realisiert, sie stellen somit keine spezifischen Probleme des Gegenstandsbereichs Textretrieval dar.

In GESINE sollen folgende Zusatzwerkzeuge, die z.T. als eigenständige Entwicklungen entstanden, integriert werden:

- Filter als relativ stabile Voreinstellungen für Recherchen, mit denen etwa die Auswahl der zu recherchierenden Datenbanken oder der Ausschluß bestimmter Informationsarten voreingestellt werden können

- Thesaurus und Klassifikation Sozialwissenschaften zur Unterstützung der Verwendung eines kontrollierten Vokabulars bei der Suche
- Rechercheverwaltung und -ablage einschließlich einer Protokollfunktion für die einzelnen Arbeitsschritte einer Sitzung
- Export-Werkzeug, welches den Übertrag der Ergebnismengen der Recherchen in andere Programme (Datenbanken oder Textverarbeitungsprogramme) in verschiedenen Formaten unterstützt
- Komponente, mit deren Hilfe eine recherchierte Menge von Dokumenten entsprechend einer inhaltlichen Gliederung sortiert werden kann. Die Einbeziehung von COGET (steht für „computergestützte Erstellung von Themendokumentationen“) bietet sich hierfür an. COGET ist eine Eigenentwicklung des IZ und wird bereits seit 1996 bei der Herstellung des Sozialwissenschaftlichen Fachinformationsdienstes soFID angewandt (cf. Riege et al.(1996))
- Möglichkeit zur Standortrecherche für gefundene Literaturnachweise via Internet mit Hilfe der ISBN/ISSN (als Teil des Programms DEUBIB zur Auswertung der Deutsche Bibliographie seit Anfang 1998 in der IZ-Bibliothek bei der Literaturbestellung eingesetzt)

Die Entwicklung der Programmteile für GESINE vollzog sich arbeitsteilig innerhalb der Projektgruppe „Informationssysteme“. Nach ausführlichen theoretischen Vorüberlegungen und konzeptionellen Arbeiten (cf. Marx et al. 1995) war im November 1996 Beginn der eigentlichen Programmierarbeit.

2.2 Grundkonzept für das Oberflächendesign

Bei der Konzeption einer elektronischen Version des Thesaurus Sozialwissenschaften war von verschiedenen Anforderungen der potentiellen Nutzer auszugehen. Zum einen war der Einsatz als unterstützendes Werkzeug bei einer Recherche mit dem Programm GESINE vorgesehen. Begriffe aus dem Thesaurus sollten in eine Recherchelogik übertragbar sein. Zum anderen sollte der Thesaurus, wie das gedruckte Werk auch, bei der inhaltlichen Erschließung von Dokumenten für die Datenbanken SOLIS und FORIS durch die Vergabe von Deskriptoren als Nachschlagwerk zur Orientierung dienen.

Folgende Design-Entscheidungen waren damit für die Konzeption der Programmoberfläche festgelegt:

- Anzeige einer alphabetische Liste aller Thesaurusbegriffe
Deskriptoren und Nichtdeskriptoren des Thesaurus Sozialwissenschaften sind im gedruckten Werk gemeinsam dargestellt. Bei vergleichbaren Lösungen (vor allem auf verschiedenen CD-ROM) hat sich eine alphabetisch sortierte Liste des kontrollierten Vokabulars als praktikabel in der Handhabung erwiesen. Die zu einem Begriff der Liste vorhandenen Treffer in der Datenbank werden bei manchen Lösungen (bspw. CD-ROM WISO) zusätzlich mit angezeigt. Das erleichtert dem Informationssuchenden bereits an dieser Stelle die Auswahl der Begriffe, da bereits vor Rechercheausführung Ergebnismengen abgeschätzt und bei der Formulierung der Anfrage zu große oder zu kleine Treffermengen ausgeschlossen werden können.
- Möglichkeit der inkrementellen Suche in der Begriffsliste
Eine inkrementelle Suche ist bei einem Umfang von ca. 10.500 Begriffen Voraussetzung für die schnelle Orientierung in dieser Liste. Um einen dem Benutzer nachvollziehbaren Suchvorgang abzubilden, erscheint ein separates Eingabefeld sinnvoll. In den Hilfswerkzeugen gängiger Software für Windows ist in der Regel auch ein separates Eingabefeld vorhanden, damit ist ein Wiedererkennungswert gegeben. Das Listenfeld soll seine Anzeige direkt nach jeder Eingabe aktualisieren, d. h. die Anzeige scrollt zum ersten Begriff der Liste, der mit dem eingegebenen Text beginnt. Andere Lösungen haben eine inkrementelle Suche direkt in der Begriffsliste realisiert. Der eingegebene Text wird dabei z.T. durch farbiges Unterlegen des Wortanfangs gekennzeichnet (in der DOS-Version der CD-ROM WISO). Diese Lösung ist m.E. suboptimal, weil Anzeige- und Eingabemöglichkeit zusammengefaßt sind. Vom Benutzer wird eine nicht selbsterklärende Handlung erwartet, nämlich die Eingabe in ein nichteditierbares Feld. Manche Anbieter reduzieren die Begriffsliste während der Eingabe auf jene Einträge, die mit dem eingegebenen Text beginnen. Das hat den Nachteil, daß Begriffe aus dem Umfeld der Liste, die einen anderen Wortanfang haben, erst nach Editieren des Eingabefeldes und nicht durch Benutzung des Scrollbalkens der Anzeige erreicht werden können. So würden bei Eingabe von „*Politik*“ zwar alle Begriffe stehenbleiben, die mit „*Politik*“ beginnen, aber Begriffe, die mit „*politische*“ anfangen, nicht mehr angezeigt, obwohl sie direkt auf „*Politik*“ folgen.

- **Anzeige aller zu einem Begriff vorhandenen Informationen**

Im gedruckten Thesaurus kann der Leser zu jedem Eintrag fast alle Verweise unmittelbar nachlesen. Bei Deskriptoren, auf die von einem Nichtdeskriptor aus verwiesen wird, ist die Informationsbeschaffung nur durch zusätzliches Blättern möglich. Beim Deskriptor „*Absatz*“ bspw. findet man „*UF Absatzweg+*“, was im Klartext bedeutet, daß der Deskriptor „*Absatz*“ in Kombination mit einem hier nicht zu erkennenden Deskriptor für den Nichtdeskriptor „*Absatzweg*“ verwendet wird. Nach Blättern findet man, daß „*Organisation+Absatz*“ für „*Absatzweg*“ benutzt werden soll. Für die Gestaltung der Anzeige sollte eine Lösung gefunden werden, die solche „benutzt für“-Verweise direkt auflöst und dem Anwender unter Umgehung des Eingabefeldes das sofortige Navigieren innerhalb der Struktur ermöglicht. Dafür bot sich eine hypertextbasierte Lösung an, bei der alle Verweise als Hyperlinks gestaltet werden. Die übrigen Anzeigefelder (Begriffsliste, Eingabefeld) aktualisieren sich im Sinne der kontextsensitiven Durchlässigkeit bei Betätigung der Links.
- **Umschalten zwischen alphabetischer und systematischer Darstellung**

Mit der Realisierung einer Möglichkeit zum „Umschalten“ zwischen der systematischen und alphabetischen Struktur wird die im Vergleich zur alphabetischen Liste weniger gebräuchliche systematische Sortierung des Thesaurus elegant einbezogen. Als Lösung empfahl sich die Verwendung sogenannter Tabpages, deutsch oft als „Karteikarte“ bezeichnet. Diese verdeutlicht auch optisch die doppelte Zuordnung eines Begriffs in einer alphabetischen Liste zum einen und einer systematischen Hierarchie zum anderen. Beim Wechsel der Tabpage wird der aktuelle Begriff beibehalten und nur die Betrachtungsperspektive (Einordnung alphabetisch oder systematisch) ändert sich.
- **Systematischer Teil des Thesaurus**

Der systematische Teil des Thesaurus ist als separates Druckerzeugnis erschienen. Dort sind zu jeder Notation in alphabetischer Reihenfolge die zugehörigen Begriffe einschließlich ihrer Verweise aufgelistet. Da diese Ordnung streng hierarchisch aufgebaut ist, bot sich für die elektronische Version eine entfaltbare Baumdarstellung an, wie sie aus dem Dateimanager bzw. Explorer von Windows bekannt ist. Damit ist für den Anwender die Einordnung klar und übersichtlich nachzuvollziehen. Bei großen Systematikgruppen (etwa Gruppe 5.6.00: Namen von Organisationen und In-

stitutionen, mit über 300 dazugehörigen Einträgen) läßt sich das Blättern naturgemäß nicht verhindern.

- **Darstellung bereits besuchter Einträge**
Die während einer Sitzung „besuchten“ Begriffe sollen nachvollziehbar, bspw. in der Reihenfolge ihres Aufrufs, in einem „history-element“ gespeichert werden. Damit soll die Navigation in einem für die aktuelle Sitzung relevanten Bestand bereits angeschauter Begriffe erleichtert werden. Zugleich wird die Gefahr des „lost in hyperspace“ gebannt. Internet-Browser besitzen eine analoge Funktionalität.
- **Exportmöglichkeit in andere Programme**
Von der Ausgangskonzeption des Thesaurus als Werkzeug für das Recherchemodul von GESINE war zunächst eine Übernahme in die Suchmaske von GESINE vorgesehen. Prinzipiell im Sinne einer allgemeinen Verwendbarkeit des Thesaurus sollte eine Schnittstelle zu anderen Programmen vorgesehen sein. Die konkrete Konzeption dafür konnte jedoch erst im Prozeß des rapid prototyping erarbeitet werden. Sie ergab sich dann aus direkten Anforderungen der Arbeitsprozesse im IZ (vergl. dazu Kapitel 3.3.4).
- **Anzeige der Treffermenge**
Bei gängigen Softwarelösungen und vielen online-Datenbanken wird als Hilfsmittel für die zu erwartende Recherchepräzision die Treffermenge angezeigt. Das hat sich als sehr hilfreich vor allem für weniger routinierte Anwender erwiesen. Im Verlauf der Softwareentwicklung ist nach Rückkopplung mit den Anwendern in einem mehrstufigen Verfahren eine vielgestaltige Trefferanzeige realisiert worden (vergl. dazu Kapitel 3.1).

Das relationale Datenmodell der IZ-Daten war bei Beginn der Entwicklungsarbeiten bereits realisiert (cf. Mutschke 1996), die thesaurusrelevanten Tabellen aus der aDIS-Datenbank nach ORACLE übernommen worden.

3 Rapid Prototyping und Anwendertests

Generell hat sich für alle Softwareentwicklungen im IZ (und darüber hinaus) das Verfahren des rapid prototyping bewährt. Durch die frühzeitige Einbeziehung der zukünftigen Anwender können so weitgehend Fehlentwicklungen minimiert und Designentscheidungen im Vorfeld beraten und gegebenen-

falls korrigiert werden (ausführlichere Beschreibung des Verfahrens mit seinen theoretischen Hintergründen cf. Riege et al. (1996), S. 7-10).

3.1 Entwicklung des Prototypen

Nach Erstellung des GESIS-weiten relationalen Datenmodells und Übertrag der thesaurusrelevanten Daten aus aDIS in eine ORACLE-Datenbank begannen im Dezember 1996 die Entwicklungsarbeiten an GESINE. Arbeitsteilig wurden einzelne Komponenten der Software mit dem Ziel realisiert, sie in GESINE integrativ einzusetzen. Für die Suchmaske von GESINE war von Beginn an eine Komponente vorgesehen, die den für die Anwenderdomäne relevanten Thesaurus Sozialwissenschaften abbildet.

Ein erster bereits funktionsfähiger Entwurf konnte im Januar 1997 einer Expertenrunde des IZ präsentiert werde.

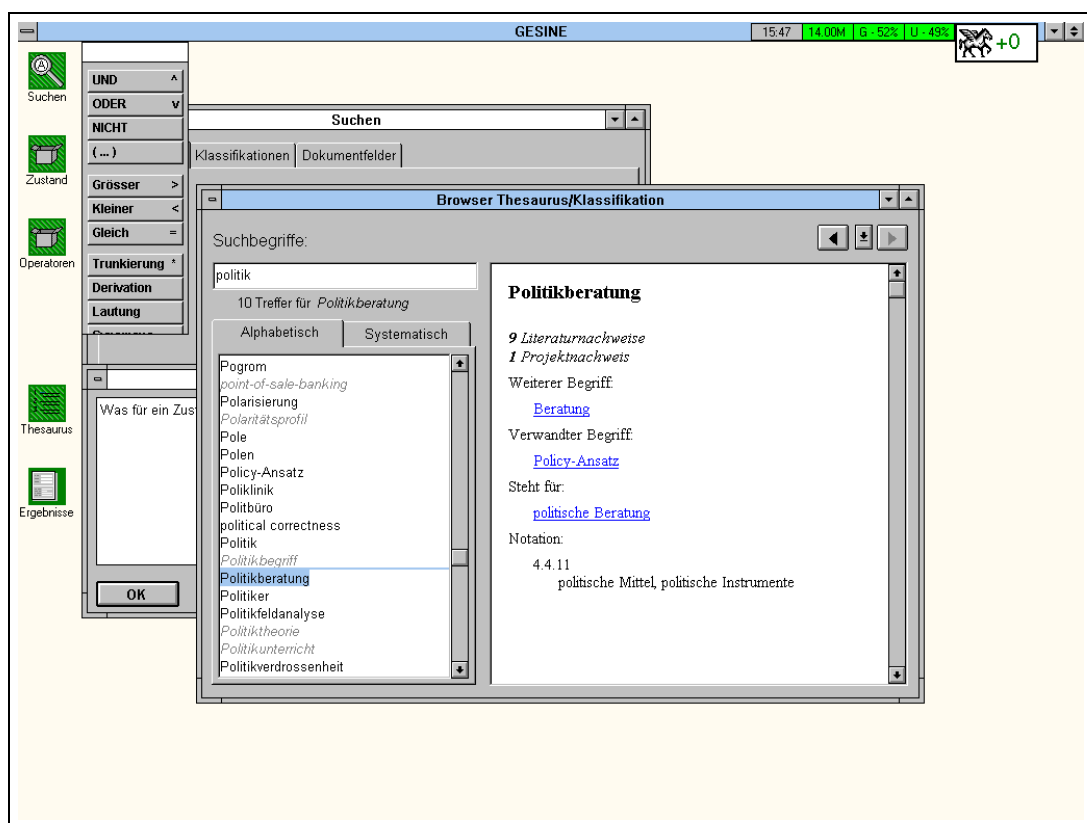


Abbildung 1: Erster Entwurf für den Thesaurus in der GESINE-Oberfläche

Die grundlegenden Elemente der Oberflächenstruktur sind hier bereits implementiert. Das Eingabefeld wurde als separates Objekt plazierte, mit einem

Ettikett („Suchbegriff:“) versehen, die Realisierung erfolgte als einzeliliges editierbares Feld. Eine zweilaschige, beschriftete Tabpage für die alphabetische und systematische Darstellung befand sich direkt darunter, um dem Anwender den Bezug zum Eingabefeld deutlich zu zeigen. Zwischen beiden Objekten war die Trefferanzeige positioniert. Den größten Platz der Oberfläche beansprucht das Anzeigefeld für die zum Listeneintrag gefundenen Informationen und Verweise, realisiert mit einem hypertextfähigen Anzeigefeld. Mit unmittelbarem optischen Bezug zum Anzeigefeld wurde der Navigationsspeicher (history-object) mit zwei Pfeiltastfeldern und einer aufklappbaren Liste (dropdown listbox) für die Ablage der besuchten Einträge der Begriffsliste angelegt.

In Expertengesprächen ergaben sich folgende Anregungen:

- Die Trefferanzeige sollte bereits während der inkrementellen Suche Treffermengen berechnen, also nach Eingabe von „*sozialw*“ folgendes anzeigen: „3467 Nachweise für *sozialw*%“. Damit könnte bei Übernahme trunzierter Begriffe in die Suchlogik eine Mengenabschätzung erfolgen. Die häufig genutzte Recherchepraxis, durch Eingabe eines rechts trunzierten Begriffs ganze Cluster von Begriffen mit gleichem Wortanfang zu erfassen, würde so dahingehend unterstützt, daß zu erwartende Treffermengen von vornherein bekannt sind.
- Nichtdeskriptoren sollten sich von Deskriptoren in ihrer farblichen Darstellung unterscheiden (ist in der Darstellung in Abbildung 1 bereits realisiert), sowohl in der Begriffsliste wie auch im Anzeigefeld.
- Die Anzeige der Notationen zu den Begriffen der Liste fehlt (ist in der Darstellung in Abbildung 1 bereits realisiert)
- Der systematische Teil des Thesaurus fehlte noch.
- Die Berechnung der Menge der Dokumente, getrennt nach den Datenbanken SOLIS und FORIS, die mit dem aktuellen Deskriptor verschlagwortet sind, wird als nützliche Zusatzinformation empfunden. In der Anzeige zu jedem Begriff ist es nach Meinung der Experten sinnvoll, nicht nur die Treffermengen für die Deskriptoren selbst auszuweisen. Für den Rechercheur sind auch Informationen über Treffermengen verwandter, näherer oder weitere Begriffe wichtig für die Entscheidung, mit welchem Begriff in der Suche gearbeitet werden soll.

- Die Experten verweisen auf fehlerhafte Abbildung von bestimmten Verweistypen (use-Verweise). Diese waren, nach Prüfung der entsprechenden Datenbankabfragen, auf falsche Interpretation des Datenmodells seitens der Entwickler zurückzuführen.
- Zum Design wurde vorgeschlagen, die Beschriftungen der Tabpages nach unten zu verlegen, das Eingabefeld zur inkrementellen Suche auf die Tabseiten zu verlegen sowie die Trefferanzeige in den freien Platz oberhalb des Anzeigefeldes zu verschieben. Ziel dieses Redesigns war es, optisch klarere Strukturen und Zugehörigkeiten zu schaffen.

Alle von den Experten vorgetragenen Wünsche und Anregungen konnten ausnahmslos realisiert werden. In einer zweiten Runde von Expertengesprächen fand der dann vorgelegte Entwurf prinzipielle Zustimmung.

Weitere Anregungen bezogen sich auf folgendes:

- Auch im systematischen Teil ist in der Baumdarstellung eine farbliche Unterscheidung zwischen Deskriptoren und Nichtdeskriptoren wünschenswert. Das war mit dem zur Verfügung stehenden Control nicht zu realisieren. Da jedoch im Anzeigefeld Nichtdeskriptoren grau angezeigt werden, ist dieses Manko in der Darstellung m.E. vertretbar.
- Der systematische Baum weist im Gliederungspunkt 5 überflüssige, leere Ebenen auf: 5.1 taucht als 2. leere Ebene auf, Einträge finden sich erst in der Ebene 3 unter 5.1.00. Alle anderen Bereiche des Systematikbaumes haben erst ab Ebene drei Einträge von Begriffen. Aus Gründen der Symmetrie des Baumes wurden die bis dato nicht existierende zweite Ebene eingefügt.
- Die Anzeige des Systematikbaumes hat eine empirisch festgestellte obere Grenze von etwa 800 Einträgen. Diese Obergrenze wurde im Test mehrfach erreicht, wenn bei den entfalteten Baumteilen zufällig Gruppen mit sehr vielen Einträgen gewählt wurden (bei den Gruppen unter der Notation 5 bspw.). Dies erwies sich als bekannte Einschränkung, die durch die Entwickler nicht zu beeinflussen war.
- Tooltips sollten soweit wie möglich implementiert werden.
- Unsauberkeiten in der Programmierung führten zu einigen Erscheinungen an der Oberfläche, die nachzuprüfen und zu beseitigen waren. So kam es

nach Anzeige eines Deskriptors zum „Nachhüpfen“ des Textes im Eingabefeld. Zuweilen funktionierten auch die Hyperlinks der Anzeige erst beim zweiten Klick. Das Trunkierungszeichen „%“ der Trefferanzeige erschien auch bei Anzeige eines eindeutigen Begriffs. Die inkrementelle Suche nach Begriffen, die mit einer Ziffer beginnen, funktionierte nicht korrekt.

Im Rahmen der Vorbereitung der GESIS insgesamt auf die Begutachtung durch den Wissenschaftsrat führte das Entwicklerteam im Frühjahr 1997 den Prototypen von GESINE einschließlich des Thesaurus vor den Belegschaften aller Mitgliedsinstitute vor. Die Diskussionen im Rahmen dieser Veranstaltungen bezogen sich weniger auf konkrete Oberflächenanforderungen und Funktionalitäten, sondern eher auf Fragen der generellen Anforderungen an Retrievalsoftware zur Suche in heterogenen Datenbeständen. Die Diskussionen bestätigten das Entwicklerteam prinzipiell in seiner Arbeitsweise.

Ebenso erfolgreich verlief die Präsentation von GESINE vor dem Wissenschaftsrat im April 1997. Im abschließenden Gutachten werden die Arbeiten wie folgt bewertet: „In der Abteilung Forschung und Entwicklung werden mit den Schwerpunkten Integrierte Informationsdarstellung und -aufbereitung sowie Verbesserung der Informationsabläufe zwei Schwerpunkte sehr kompetent bearbeitet. ... Das Projekt GESINE bietet im Rahmen der integrierten sozialwissenschaftlichen Informationssysteme interessante Möglichkeiten, heteogene Daten zu kombinieren; Es erscheint in seiner Konzeption benutzergerecht und könnte die Akzeptanz von Informationsdiensten generell verbessern. Eine engere Zusammenarbeit und Abstimmung zwischen informationstechnischer Forschung und Entwicklung und sozialwissenschaftlichen Nutzern ist jedoch erforderlich, damit die vorgesehenen und wünschenswerten technologischen Innovationen eine breitere Wirkung erfahren können.“ (cf. Wissenschaftspolitische Stellungnahme des Wissenschaftsrates zur GESIS , S. 47).

3.2 Anwendertest Mai 1997

Nach Beseitigung der im ersten Anwendertest festgestellten Unzulänglichkeiten konnte im Mai 1997 die Freigabe für einen institutsweiten Einsatz erfolgen. Dazu wurde der Thesaurus von GESINE abgekoppelt und als eigenständiges Produkt im Netzwerk des IZ zur Verfügung gestellt. Die Anwender erhielten Zugang zur ORACLE-Datenbank. Am Test selbst beteiligten sich

14 Mitarbeiterinnen und Mitarbeiter, in deren Arbeitsgebiet die Benutzung des Thesaurus Sozialwissenschaften fällt, die also an der intellektuellen Erschließung von Dokumenten (Vergabe von Deskriptoren) arbeiten bzw. den Thesaurus als Nachschlagwerk für Auftragsrecherchen nutzen.

An alle Testpersonen wurde ein sechsseitiger Fragebogen verteilt. Ziel war es zu testen, wie mit der Software gearbeitet wird, welche der angebotenen Funktionalitäten in welcher Weise genutzt werden. Die Probanden waren aufgefordert, die einzelnen Funktionen zu beurteilen, deren individuelle Verwendung zu beschreiben sowie zu begründen, warum sie diese anwenden bzw. warum nicht. Fünf detailliert ausgefüllte Fragebögen standen zur abschließenden Auswertung zur Verfügung.

Die auswertbaren Fragebögen lieferten Beurteilungen und Einschätzungen aus verschiedenen, aufgabenspezifischen Sichtweisen auf die Software. Es konnten die Beurteilungen aus der Sicht eines Rechercheurs (Versuchspersonen 1 und 5), der formalen Kontrolle und Indexierung (Versuchsperson 2), der inhaltlichen Erschließung und Indexierung (Versuchspersonen 3, 4 und 5) in die Auswertung einfließen. Damit waren alle relevanten Arbeitsbereiche des IZ vertreten.

Die Fragen zu den einzelnen Bereichen des Thesaurus-Browsers wurden wie folgt beantwortet.

Allgemeine Funktionen

- Der Thesaurus-Browser ist angenehmer in der Anwendung als das Blättern in der Druckversion (VP 2).
- Der systematische Teil wurde selten verwendet (VP 2, VP 4).
- Die alphabetische Liste ist eine gute Hilfe bei der Ermittlung der Zulässigkeit von Begriffen (VP 4).

Suche von Deskriptoren

- Das Übertragen einzelner Deskriptoren (mit der Tasten-Kopierfunktion <STRG>+<C>, dann einfügen) in eine Upload-Datei zur Recherchevorbereitung ist hilfreich (VP 1).

- Das Markieren von Begriffen ist umständlich. Wenn der Cursor im Schreibfeld steht, dort aber nur der Wortanfang getippt wurde, muß erst der ausgewählte Deskriptor angeklickt werden, um ihn markieren zu können (VP 5).
- Die Trefferanzeige insgesamt sowie die Anzeige der Vergabehäufigkeit des aktuellen Deskriptors ist „schön und hilfreich“ bei der Recherchevorbereitung (VP 1, VP 4).
- Die „history-Funktion“ wird selten eingesetzt (VP 3).
- Zur Klappliste mit den besuchten Begriffen: „Solange eine Sammelfunktion fehlt, arbeite ich Deskriptor um Deskriptor ab. Das geschieht mehr oder weniger systematisch je nach Thema (und Kompetenz meinerseits). Die Rückkehr zu an sich geklärten Begriffen geschieht selten. Die von der Dauer der Sitzung etc. abhängige Frage ist, ob sie über die Klappliste beschleunigt wird i.V. zur Neueingabe im Suchfeld. Fazit: M.E. eignet sich diese Funktion weniger für das Indexieren als Produktionsvorgang. Vermutlich aber für Recherche-Unterstützung.“ (VP 4)
- Das Aufklappen der systematischen Gliederungspunkte der Baumstruktur stellt eine plausible Lösung dar (VP 4).
- Die Trefferanzeige arbeitet unkorrekt. Nach Eingabe von „*Wissenschafts*“ wird „*Wissenschaft*“ angezeigt, die Trefferanzeige jedoch für „*wissenschafts%*“ ausgewiesen (VP 5).
- Durch die intern realisierte Logik des Aufbaus der Anzeige kommt es häufig zu unangenehmen Effekten. Bei der Eingabe von „*forschungspolit*“ springt der Browser zu „*Forschungspolitik*“ und dann wieder zurück zu „*Forschung*“. Der Effekt trat auch bei anderen Begriffen auf („*alternativ*“ wird zu „*Alter*“)(VP 5).

Strategie der Deskriptorensuche

Dieser Fragenkomplex wurde von keinem Probanden ausgefüllt. Er sollte bei weiteren Tests nur durch teilnehmende Beobachtung erhoben werden (VP 4).

Orientierung in der Hypertextdarstellung

- Bei der Hypertextdarstellung ist es hilfreich, alle Informationen einschließlich der Kombinationsverweise auf einen Blick zu haben (VP 1).
- Die Darstellung der Kombinationsverweise ist für jemanden, der nicht täglich mit dem Thesaurus arbeitet, wichtig (VP 2).
- Bei der formalen Kontrolle von Partnerlieferungen ist der Thesaurus hilfreich für die Suche nach einem „besseren“ Begriff, falls ein vergebener Deskriptor „fragwürdig“ ist (VP 2).
- Das Umschalten auf den systematischen Teil durch einen Klick auf die Notation wird selten eingesetzt (VP 3).

Fehlende Funktionalitäten und Fehler

- Eine Übersetzung ins Englische zur Vorbereitung von Recherchen in internationalen Datenbanken fehlt, französisch und spanisch wäre auch wünschenswert (VP 1).
- Ein „Copy-Knopf“ für die Windows-Zwischenablage fehlt (VP 3).
- Die Thesaurus-Merkmale „Register-Haupt- und Nebeneintrag“ fehlen (VP 4, VP 5).
- Eine Sammel- und Kopierfunktion fehlt. Die Sammeldatei sollte eine Edierfunktion haben. Wünschenswert wäre auch das „Kopieren von Begriffskombinationen aus der Gesamtanzeige“. Die Übertragung könnte per Button erfolgen (VP 4).
- Eine „Löschtaaste“ für das Suchfeld würde die Eingabe beschleunigen (VP 4).
- Eine Kommentiermöglichkeit zu den Einträgen im Thesaurus für persönliche oder allgemeine Bemerkungen fehlt (VP 4).

Handhabung der Software

Diese offene Frage wurde von keiner Versuchsperson ausgefüllt. Aufgrund umfangreicher Schulungsmaßnahmen zu Windows im Institut bestanden kei-

ne Schwierigkeiten mit der Bedienung. Ebenso bereitet der Umgang mit Hypertexten keine Schwierigkeiten.

Vergleich zwischen bisheriger Arbeitsweise und der mit dem Thesaurus-Browser

- Die Benutzung des Thesaurus spart kaum Arbeitszeit ein, da die Thesaurusbegriffe weitgehend bekannt sind. (VP 1).
- Die elektronische Version ist „angenehmer und schneller als Blättern in Druckversion“ (VP 2).
- Ein Nachteil der elektronischen Version besteht darin, daß noch mehr Bildschirmarbeit nötig ist mit den entsprechenden Folgen, wie höhere Augenbelastung und weniger Änderungen in der Sitzposition (VP 3).
- Der Vorteil besteht darin, daß die Übernahme von Deskriptoren (via Tastenkombination und Windows-Zwischenablage) möglich ist, was insbesondere bei längeren Begriffen arbeitserleichternd ist, da es Schreibarbeit einspart (VP 3).
- Die Baumdarstellung des systematischen Teil ermöglicht eine bessere Einsicht in dessen Struktur (VP 3).
- „In der Druckversion kommentiere ich fortlaufend, füge ein, was bei der nächsten Thesaurus-Revision einzubringen ist.“ Das geht in der elektronischen Version nicht (VP 4).
- Der Wechsel zwischen Papiervorlage auf dem Schreibtisch (zu indexierendes Dokument, z.B. ein FORIS-Erhebungsbogen), dem Thesaurus-Browser auf dem Bildschirm als Nachschlagwerk, der folgende Kopiervorgang, der Wechsel in die aDIS-Anwendung, das Einfügen aus der Zwischenablage und zurück zur Papiervorlage ist anstrengend, erfordert starke Konzentration (VP 4).

Technische Aspekte

- Das Programm „verhaspelt“ sich bei der Eingabe, beim Überschreiben, bei Wortstämmen, die auf gültige Eintragungen treffen (VP 3).
- Die Wartezeit beim Aufruf des Thesaurus ist zu lang (VP 4).

- Zeitweise kommt es zu „unerträglich langsamen Reaktionen“ bei der Sucheingabe (VP 4).
- Das rote Ausrufezeichen bei Verweisen wird nicht abgebildet (VP 4).

Weitere Bemerkungen, Anregungen, Wünsche

- Ein getrennter und gleichzeitiger Zugriff auf alphabetische und systematische Liste könnte ein Indexieren mit der alphabetischen und ein paralleles Nachblättern in der systematischen Liste ermöglichen, ohne bspw. die alphabetische Darstellung zu verändern (VP 4).
- Zur Systemkonzeption: „Wenn das System so funktionieren würde, wie es sollte, wäre es ... toll und eine große Arbeitserleichterung.“ (VP 5)

Der erste Testeinsatz des Thesaurus-Browsers verlief insgesamt erfolgreich. Alle Versuchspersonen kamen mit der Oberfläche grundsätzlich gut zurecht. Die Art der aus den Protokollen und begleitenden Gesprächen gewonnenen Kritik ist durchweg konstruktiv. Sie weist lediglich auf Schwachstellen oder zusätzlich gewünschte Funktionalitäten hin. Die generelle Eignung für die Arbeitsgebiete „inhaltliche Erschließung von Dokumenten“ sowie „Vorbereitung von Recherchen“ wurde bestätigt. Die Hinweise und Anregungen wurden in der weiteren Entwicklungsarbeit zu großen Teilen berücksichtigt.

3.3 Weiterentwicklung bis Dezember 1997

3.3.1 Änderungswünsche, Verbesserungen und Korrekturen

Nach dem ersten Anwendertest konnten eine Reihe von Anregungen realisiert werden.

Die Trefferanzeige oberhalb des Anzeigefensters aktualisierte sich in der ersten Version nach jeder Eingabe sofort. Messungen ergaben, daß diese Berechnungen die Performanz der Applikation erheblich beeinträchtigten. Besonders bei versierten Benutzern der Tastatur vollzog sich die Eingabe deutlich schneller als die Berechnungen für die Trefferanzeige. Dadurch „verhaspelte“ sich die Anzeige; erst deutlich nach Ende der Eingabe vollzog sich ein schrittweises Aktualisieren der Trefferzahl. Eine vorläufig befriedigende Lösung konnte gefunden werden, indem hier eine Zeitverzögerung

eingeschaltet wurde. Diese setzte den Berechnungsmodus für die Trefferanzeige erst in Gang, wenn in einer bestimmten Zeitspanne keine Tastendrucke zu verzeichnen waren. Durch Messungen hat sich ein Wert von 0,3 Sekunden als ausreichend erwiesen. Schnellschreiber schaffen es, mit dieser Zeitverzögerung die Anzeige erst nach Abschluß der Eingabe auszulösen. Für langsame Schreiber war dies ohnehin kein Problem.

Die im Anwendertest angeregte Einbeziehung der Verwendungsmöglichkeit eines Begriffs als Registerhaupt- oder -nebeneintrag bei der Indexierung von Forschungsprojekten wurde wegen der geplanten Einstellung der gedruckten Version der „Forschungsarbeiten in den Sozialwissenschaften“ nicht realisiert. Registerbegriffe wurden bis dato ausschließlich zur Erzeugung des mehrstufigen Registers dieser Publikation sowie reiner Forschungsdokumentationen verwendet.

Eine Kommentarfunktion für persönliche oder allgemein zugängliche Bemerkungen zu einzelnen Begriffen des Thesaurus, wie sie von einem Anwender angeregt wurde, kam hier nicht zur Realisierung. Der dahinterliegende Gedanke, mit Hilfe eines zusätzlichen editierbaren Feldes bspw. Notizen zur Verwendung von Begriffen zu speichern oder die notwendige Einführung neuer engerer Begriffe zu vermerken, fand bei anderen Anwendern keine Unterstützung.

Die inkrementelle Suche arbeitet bei einigen Begriffen mit Umlauten nicht korrekt, d. h. die Eingabe und die Listenanzeige synchronisieren sich nicht parallel. Ursache dafür ist die im Programm realisierte Logik der binären Suche, die auf dem Vergleich zweier Zeichenketten basiert. In der Begriffsliste werden die Umlaute nach ihren dazugehörigen Nichtumlauten einsortiert, wenn die Begriffe sonst identisch sind. Der Eintrag „Wahl“ stünde also vor dem (fiktiven) Eintrag „Wähl“. Bei nichtidentischen Strings wird der Umlaut wie der Nichtumlaut einsortiert, also steht „Müll“ nach „muckracking“. Problematisch bei der binären Suche ist der Stringvergleich, d.h. die Entscheidung, ob die im Eingabefeld stehende Zeichenkette vor oder nach dem angesprochenen Begriff in der Liste steht. Beispiel: Bei der Eingabe von „ü“ sollte in der Liste zum ersten mit „ü“ beginnenden Begriff („Überalterung“) gescrollt werden. Stattdessen bleibt der Browser beim gerade aktuellen Begriff stehen. Der Such- und Vergleichsalgorithmus muß für solche Fälle noch so überarbeitet werden, daß alle Zeichenketten mit Umlauten korrekt behandelt werden.

3.3.2 Klassifikationsbrowser

Im Mai 1997 begannen die Entwicklungsarbeiten an einer elektronischen Version der Klassifikation Sozialwissenschaften. Da sich Thesaurus und Klassifikation im grundsätzlichen Aufbau stark ähnlich sind, konnte das Gestaltungsprinzip des Thesaurus im wesentlichen übernommen werden. Wie der Thesaurus kann die Klassifikation alphabetische oder in Form einer hierarchischen Baumstruktur dargestellt werden. Suchmöglichkeiten nach Klassifikationsnummern bzw. in der Index-Liste sind erforderlich; Verweise auf ihre Verwendungshäufigkeit in SOLIS und FORIS als Haupt- und Nebenklassifikation sind wünschenswert. Damit bot sich das durch den Anwendertest des Thesaurus-Browsers bestätigte Design zur Übernahme an. Zwei Tabpages mit einem Eingabefeld für die inkrementelle Suche sowie ein Anzeigefeld, welches die Begriffsstruktur und alle Verweise als Hypertexte darstellt, bilden auch hier das Grundkonzept für die Gestaltung der Oberfläche.

Alle Daten zur Klassifikation lagen in maschinenlesbarer, aber nicht datenbankbasierter Form (als Textdateien) vor. Sie konnten mit relativ geringem Aufwand in das vorhandene relationale Datenmodell überführt werden.

Im Juli 1997 konnte der bisherige Thesaurus-Browser um die Klassifikation erweitert in den institutsweiten Anwendertest überführt werden. Für die Klassifikation wurde ein separates Icon auf der Oberfläche positioniert. Beim Start der Anwendung erschien nun nur noch die Werkzeugleiste mit den Icons für Thesaurus und Klassifikation. Auf eine Ausgabe von Fragebögen wurde diesmal verzichtet, da die zu erwartende Benutzung auf wenige Personen eingeschränkt war. Die beim Thesaurus schon im ersten Test sichtbar und wirksam gewordenen Arbeitserleichterungen waren für die Klassifikation nicht zu erwarten. Die 158 verschiedenen Klassen sind auch für ungeübtere Anwender leichter zu überblicken als 10.500 Thesauruseinträge. Ebenso ist das Tippen von wenigen fünfstelligen Nummern pro Inhaltserschließung nicht so fehlerbehaftet, wie die Eingabe von mehreren, zum Teil langen Deskriptoren.

Im Verlauf der Entwicklungsarbeiten deutete sich eine für die bisherige Printversion nur schwer realisierbare Erweiterungsmöglichkeit an. Der bis dato vorhandene Index, bestehend aus den Klassifikationsbegriffen selbst und einigen Fachbegriffen, die als Verweise auf dafür zu verwendende Klassifikationen aufgenommen wurden, war in seiner Form als Textdatei nur schwer zu pflegen bzw. zu ergänzen. Die Klassifikation als Grundschema

sollte von größerer Beständigkeit sein. Neu entstehende Fachgebiete oder Teildisziplinen werden so in der Regel nicht direkt abbildbar sein. Deshalb ist es wichtig, eine möglichst aktuelle Verweisstruktur aufzubauen, die von solch neuen Begriffen auf dafür zu verwendende Klassifikationen verweist. Mit der Überführung in das Datenmodell sind Ergänzungen der Verweisstrukturen nun einfacher möglich.

3.3.3 Umstellung auf Windows 95

Im November 1997 erfolgte nach Einführung von Windows 95 im Institut eine Anpassung des Thesaurus- und Klassifikationsbrowsers. Bedingt durch die Verwendung einer externen Programmbibliothek für das Entwicklungswerkzeug PowerBuilder, die nicht vollständig kompatibel mit Windows 95 war, gab es an der Anwendungsoberfläche Probleme mit der Darstellung der Icons. Gelöst werden konnte dies Problem mit der Einfügung eines neuen Programmfensters, auf dem die Icons plaziert wurden. An der Überführung der für diese Unzulänglichkeit ursächlichen Programmbibliothek in eine Windows 95 -kompatible Form wird gearbeitet.

3.3.4 Sammelbeutel

Die aus erstem Anwendertest gewünschte Sammelfunktion bzw. Exportmöglichkeit für Deskriptoren Mengen stellte einen weiteren Schwerpunkt in der Entwicklungsarbeit dar. Das Werkzeug „Thesaurus“ hatte sich in seiner Handhabung bewährt und war von den Anwendern als Arbeitsinstrument akzeptiert und eingesetzt worden.

Damit erhielt die im Verlauf des ersten Anwendertests formulierte Forderung nach Entwicklung einer einfach handhabbaren Übernahmefunktion für Deskriptoren in die Terminalemulation EMF[®]97-WIN, die am IZ bei der Inhaltserschließung von Dokumenten als Kommunikator mit der Datenbank eingesetzt wird, eine wachsende Priorität. Der manuelle und nur einzeln zu vollziehende Kopiervorgang über die Windows-Zwischenablage war zu zeitaufwendig und zu anfällig für Bedienfehler. Optimal für die Unterstützung der Arbeitsabläufe wäre das Sammeln von Deskriptoren für ein Dokument und die Übernahme „per Knopfdruck“ in die entsprechende Maske der Terminalemulation (vergl. dazu Kapitel 3.2).

Im Rahmen der Gesamtkonzeption für GESINE waren sowohl der Thesaurus als auch die Klassifikation als Werkzeuge für die Vorbereitung einer Suchan-

frage konzipiert. Von daher war ohnehin eine Übernahmefunktion zu realisieren. Die Entwicklung einer auf die speziellen Parameter der Terminalemulation EMF[®]97-WIN zugeschnittenen Lösung für die institutsinterne Anwendung bot sich also an. Da mehrere Mitarbeiter ständig und zahlreiche andere Mitarbeiter sporadisch mit der Indexierung von FORIS- und SOLIS-Dokumenten beschäftigt sind (im IZ werden pro Jahr etwa 14.000 Dokumente selbst verschlagwortet), ist eine Arbeitserleichterung und -beschleunigung in diesem Bereich auf längerfristige Sicht durchaus sinnvoll.

In ausführlichen Gesprächen mit den Anwendern und dem aDIS-Systemverwalter wurden die Randbedingungen für die Realisierung des Sammelbeutels beraten. Das Konzept für die Lösung hatte danach folgenden Arbeitsablauf, der sich durch den Arbeitsinhalt zwangsläufig ergibt, in geeigneter Weise zu unterstützen:

1. Ein SOLIS- oder FORIS-Dokument wird zur Indexierung ausgewählt.
2. Die geeigneten Deskriptoren werden nacheinander im Thesaurus-Browser ausgewählt.
3. Die Übernahme in einen Sammelbeutel erfolgt per Knopf/Button.
4. Idealerweise findet eine Dublettenprüfung innerhalb der Übernahmemenge statt.
5. Per Knopfdruck werden alle Einträge der Liste des Sammelbeutels in die aDIS-Maske übertragen. Dazu wird in der Terminalemulation zur richtigen Maske gewechselt. Die freien Stellen werden der Reihe nach aufgefüllt. Diese Einschränkung muß beachtet werden, weil bereits Deskriptoren vergeben sein können (bspw. durch direkte Eingabe ohne Benutzung des elektronischen Thesaurus oder durch vorhergehenden Einsatz der Textanalysekomponente).
6. Nach erfolgreicher Übernahme wird die Bearbeitung des Dokuments abgeschlossen (Systemprüfung und Speicherung in den kontrollierten Datenbestand) und ein neues Dokument geladen.
7. Die Einträge im Sammelbeutel sollten dort einzeln oder insgesamt löschtbar sein.
8. Nach dem Löschen aller Einträge des Sammelbeutels kann in gleicher Weise für ein anderes Dokument verfahren werden.

Die Verbindung zwischen Thesaurus-Browser und Terminalemulation wurde über eine DDE-Schnittstelle realisiert. Alle oben genannten Arbeitsabläufe

konnten in der beschriebenen Weise für den ersten Prototypen realisiert werden.

Im Dezember 1997 konnte der Sammelbeutel in den Anwendertest überführt werden.

3.3.5 Mehrsprachigkeit

Parallel zur Entwicklung des Sammelbeutels wurde an der Realisierung einer mehrsprachigen Version des Thesaurus-Browsers gearbeitet. Mit der Implementation einer mehrsprachigen Version sollten verschiedene Anforderungen verifiziert werden. Zum einen konnte überprüft werden, ob die Mehrsprachigkeit ohne Brüche im Design zu realisieren war, d.h. ob der Designentwurf sich als hinreichend stabil erwies. Zum anderen waren noch Anforderungen aus dem ersten Anwendertest offen geblieben. Insbesondere die Rechercheure hatten eine zumindest englische Version des Thesaurus-Browsers angeregt, um Upload-Dateien für Suchanfragen in internationalen Datenbanken vorbereiten zu können.

Mit den im IZ verfügbaren Übersetzungen des deutschsprachigen Thesaurus ins englische und russische waren alle nötigen Daten maschinenlesbar vorhanden. Um mögliche Probleme mit der Darstellung kyrillischer Zeichen (in Zusammenhang mit den in der Applikation verwendeten controls) zu umgehen, war für die erste prototypische Lösung der Mehrsprachigkeit die englische Version zu Realisierung ausgewählt worden. Zu jedem deutschen Begriff des Thesaurus existiert eine eindeutige englische Übersetzung, zu jedem englischen Begriff, bis auf 84 Ausnahmen mit Mehrfachübersetzungen, ebenfalls. So gibt es bspw. für „demand“ die deutschen Übersetzungen „Bedarf“ und „Nachfrage“. Begriffsrelationen, wie *related terms* oder *narrower terms*, sind in englisch nicht direkt vorhanden, sondern werden über die deutsche Übersetzung der Begriffe berechnet und in die entsprechenden Tabellen des Datenmodells eingefügt. Dieses Verfahren ist sinnvoll, weil damit Änderungen in den Relationen nur an einer Stelle umgesetzt werden müssen.

Im Design des Thesaurus-Browsers waren einige Änderungen notwendig. Oberhalb des Anzeigefeldes wurde in Form einer Klappliste eine Möglichkeit zur Sprachauswahl eingefügt. Zur optisch besseren Abgrenzung von der Anzeige, der Trefferzählung und der history-Funktion wurde zusätzlich eine Linie unter der Listbox plaziert. Mit der Änderung der Sprachauswahl wechselt die gesamte Anzeige des Thesaurus. Zusätzlich als Link eingefügt wurde im

Anzeigefeld die Übersetzung in die jeweils andere Sprache. Damit war auch über Klick auf die Übersetzung das Umschalten zur anderen Sprache möglich. Der Designentwurf erwies sich, bis auf die obigen unumgänglichen Änderungen, als robust. Verschiedene Sprachen können auf diese Weise implementiert werden.

Die englische Version konnte im Februar 1998 für die Anwender freigegeben werden (vergl. Abbildung 4)

3.4 Anwendertest Februar 1998

Im Verlauf dieses Anwendertests traten die folgenden Probleme mit dem Sammelbeutel auf. Ferner konnten zahlreiche Anregungen umgesetzt werden, die im wesentlichen der Verbesserung der Unterstützung der Arbeitsabläufe dienen.

- Die zur Übernahme benutzte DDE-Schnittstelle zur Terminalemulation setzte die Umlaute nicht korrekt um. Die PowerBuilder-Applikation benutzte den ANSI-Code, die Terminalemulation den ASCII-Code. Um dieses Problem zu umgehen, löste der Sammelbeutel vor Beginn der Übernahme alle Umlaute auf, d.h. aus „Ä“ wurde „AE“ etc. Die aDIS-interne Prüfung übersetzte solche Zeichenkombinationen dann korrekt zurück.
- Der Sammelbeutel fand nicht in jedem Falle die Maske „Deskriptoren“. Als Ursache dafür konnte die nicht konstante Reihenfolge der Masken in aDIS ausgemacht werden. So werden nach Aufruf des Korrekturmodus alle korrekt ausgefüllten Masken aus der Zählung genommen, was nach Anwendung der Textanalyse häufig die Maske „Deskriptoren“ betraf. Ein anderer, in jedem Modus funktionierender Weg der Maskensuche mußte daraufhin gewählt werden.
- Das Löschen aller Einträge im Sammelbeutel sollte optional voreinstellbar sein. Zu diesem Zweck wurde eine checkbox *„Fenster nach Übernahme leeren“* zusätzlich eingefügt.
- Die Einträge im Sammelbeutel sollten dort editierbar sein. Das ist zuweilen schneller, als das Löschen eines Eintrages und das erneute Suchen und Übertragen (z.B. *„Altenpfleger“* kann schnell zu *„Altenpflege“* korrigiert werden). Eine zusätzlich farbig gekennzeichnete Eingabezeile könnte zum

direkten Eintragen kurzer Deskriptoren (z.B. „NRW“, „ZDF“) sinnvoll sein, was ebenfalls Zeitgewinn brächte. Beides wurde realisiert.

- Nach Betätigung des Buttons „Übernahme“ wäre es günstig, den Eintrag im Eingabefeld zu markieren, damit ohne separaten Löschvorgang die Suche nach einem neuen Begriff ohne Zeitverzug beginnen kann (Eingabe überschreibt Markierung). Damit konnte eine weitere Beschleunigung realisiert werden.
- Um bei der Eingabe eines Begriffes den Wechsel von Tastaturbedienung zur Mausbedienung zur Betätigung des Buttons „Übernahme“ zu vermeiden, wurde die alternative Tastenkombination <ALT>+<n> zugelassen.
- Die Anwender machten darauf aufmerksam, daß die unterschiedlichen Schriftgrößen bei Thesaurus und Terminalemulation stören. Da bei der Terminalemulation über eine Konfiguration eine etwa gleichgroße Schrift und damit eine Entlastung der Augen erreicht werden kann, konnte ein diesbezüglicher Hinweis an alle Anwender das Problem beseitigen.
- Eine weitere Beschleunigung der Arbeit kann erreicht werden, wenn nach dem Leeren des Sammelbeutels sich auch das Eingabefeld leeren und der Cursor dorthin dorthin gesetzt würde. Auch diese Anregung konnte umgesetzt werden.
- Nach jedem Eintippen in das Eingabefeld wechselt die Anzeige in der Deskriptorenliste jeweils zum ersten Begriff, der mit dem eingegebenen Text beginnt. Wenn der eingegebene Text mit einem Begriff der Liste identisch ist oder es keinen nachfolgenden Begriff mit dem gleichen Anfang gibt, wird im Anzeigefeld die dazugehörige Struktur aufgebaut. Die Eingabe von „un“ führt zur Anzeige des Deskriptors „UN“. Die Eingabe von „abk“ führt zur Anzeige des einzigen Begriffes mit diesem Wortanfang, nämlich „Abkommen“. In solch einem Fall wird der eingetippte Text im Eingabefeld automatisch um den noch fehlenden Teil ergänzt. Um diese Ergänzung kenntlich zu machen wird der Textteil farbig unterlegt. Das erleichtert dem Anwender das Erkennen der von ihm nicht eingegebenen Zeichen.

Im Testzeitraum konnte mehrmals durch Optimierungsarbeiten im Programmcode eine z.T. gravierende Beschleunigung in der Anzeige erreicht

werden. Das betraf insbesondere komplexere Datenbankabfragen über mehrere Tabellen. Testmessungen ergaben für die Zeit vom Klick auf den Eintrag in der Thesaurus-Liste bis zum Ende des Aufbaus im Anzeigefeld folgendes:

Begriff	Zeit vor Optimierung in Sekunden	Zeit nach Optimierung in Sekunden
Abfall	1,6	0,4
Abkommen	0,4	0,4
Entwicklungsland	3,8	0,6
Entwicklungsplanung	2,6	0,4
ÖTV	1,7	0,4
Theorie	4,9	0,6

Tabelle 1: Optimierungsgewinn

Besonders bei Begriffen mit langen Verweislisten, wie *Theorie* oder *Entwicklungsland*, sind die Verkürzungen deutlich spürbar. Dahingegen wird der Anwender den Zeitgewinn bei „normalen“ Begriffen kaum bemerken. Die Versuche haben auch gezeigt, daß eine Zeit von 0,4 Sekunden für den Aufbau der Anzeige offensichtlich nicht zu unterschreiten ist.

4 Beschreibung Stand März 1998

Der Browser für den deutschen und englischen Thesaurus und die Klassifikation Sozialwissenschaften läuft seit Februar 1998 im institutsweiten Einsatz. Für die Software wurde eine Kurzanleitung mit einer Produktbeschreibung im LAN des IZ zur Verfügung gestellt, die die Basis für die folgende Produktbeschreibung ist.

4.1 Starten und Beenden des Programms

Mit einem Doppelklick auf das GESINE-Symbol öffnen sich Thesaurus und Klassifikation. Eine Verbindung zur Datenbank wird aufgebaut. Im Vordergrund erscheint das Hauptfenster in Vollbildgröße mit drei Icons am linken Bildrand für die einzelnen Fenster der Anwendung.

Beendet wird das Programm über das Systemmenü in der Titelleiste oder mit der Tastenkombination <Alt>+<F4>.

4.2 Browser-Hauptfenster

Im Hauptfenster befinden sich am linken Bildschirmrand drei Icons. Diese repräsentieren die einzelnen Anwendungsfenster:



Thesaurus



Klassifikation



Sammelbeutel

Alle Anwendungsfenster werden durch Einfachklick auf das entsprechende Symbol geöffnet.

4.3 Thesaurus

Beim Öffnen des Thesaurus wird standardmäßig nur der deutschsprachige geladen. Ein Popup-Fenster wird während des Ladevorgangs eingeblendet. Das Laden selbst dauert in Abhängigkeit von der Belastung des Datenbankservers ca. 10 Sekunden. Mit dem Thesaurus wird gleichzeitig das Sammelbeutel-Fenster geöffnet.

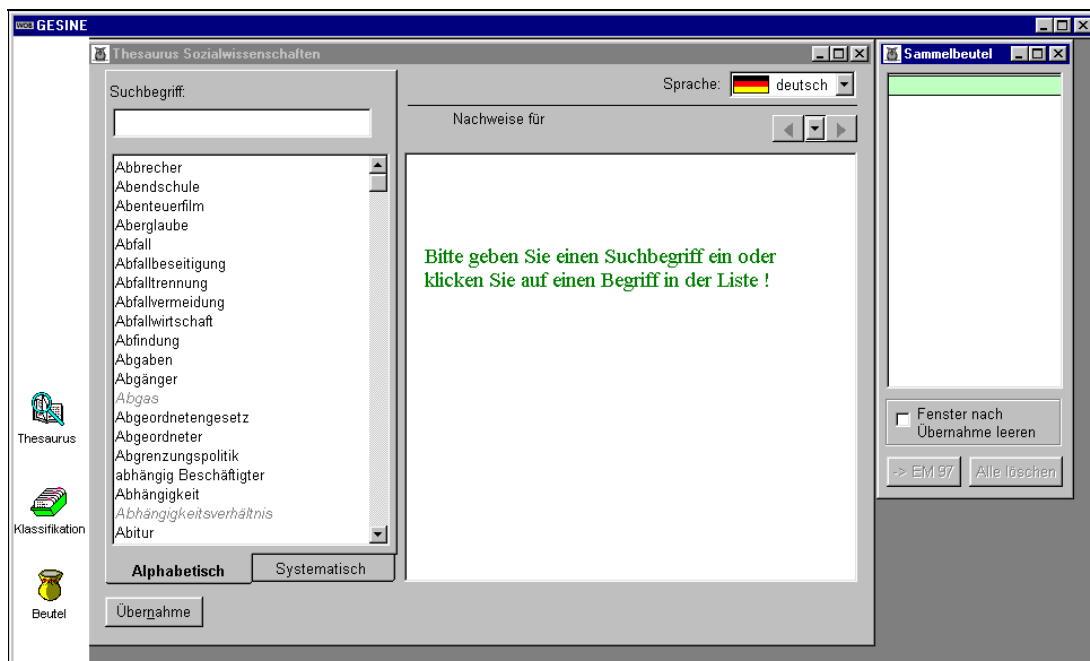


Abbildung 2: Thesaurus und Sammelbeutel im Ausgangszustand

4.3.1 Sprachauswahl

Mit der Listbox im oberen Teil des Thesaurus-Fensters kann die Sprachvariante ausgewählt werden. Derzeit stehen die deutsche und die englische zur Verfügung, die russische ist in Vorbereitung. Das erstmalige Laden jeder Sprachvariante in einer Sitzung wird durch ein Popup-Fenster angezeigt und dauert ca. 10 Sekunden. Das Umschalten zwischen den Sprachvarianten kann in jeder Arbeitsphase erfolgen. Es wird dann zu dem im Anzeigefeld dargestellten Begriff die jeweils anderssprachige Struktur aufgebaut. Gleichzeitig schalten sowohl die alphabetische Liste als auch der Systematik-Baum in die andere Sprache um. Der Umschaltvorgang dauert etwa 3 Sekunden.

4.3.2 Eingabefeld für die Suchbegriffe

Das Eingabefeld für Suchbegriffe ermöglicht die inkrementelle Suche nach Begriffen im alphabetischen und systematischen Thesaurus. Nach jeder Eingabe wechselt die Anzeige in der Deskriptorenliste zu dem ersten Begriff, der mit dem eingegebenen Text beginnt. Wenn die eingegebene Zeichenfolge mit einem Begriff der Liste übereinstimmt, wird im Anzeigefeld die dazugehörige Struktur aufgebaut (Eingabe von „un“ führt zur Anzeige von „UN“). Falls die eingegebene Zeichenfolge eindeutig auf einen Begriff der Liste verweist, wird dieser angezeigt, die fehlenden Zeichen werden ergänzt und farbig gekennzeichnet (Eingabe von „abk“ führt zum eindeutigen Begriff „Abkommen“). Das Eingabefeld kann neben der in Windows üblichen Weise auch mit der Tastenkombination <STRG>+<Y> geleert werden, unabhängig davon, an welcher Position im Eingabefeld sich der Cursor befindet.

4.3.3 Begriffsliste: Tabpage „Alphabetisch“

In alphabetischer Reihenfolge (DIN-Sortierung) sind Deskriptoren und Nichtdeskriptoren aufgelistet. Nichtdeskriptoren sind grau und in kursiver Schrift dargestellt. Innerhalb der Liste kann man sich zeilenweise mit den Pfeiltasten, der Bild↑ -Taste, der Bild↓ -Taste oder mit dem Rollbalken am rechten Rand bewegen. Die aktuelle Zeile ist blau unterlegt. Beim Wechsel zur Tabpage „Systematisch“ wird der aktuelle Begriff in seiner Zuordnung zur Systematik dargestellt, d.h. die Baumstruktur der systematischen Darstellung entfaltet sich an der entsprechenden Stelle und der aktuelle Begriff wird blau unterlegt. Das Anzeigefeld des aktuellen Begriffs bleibt dabei erhalten.

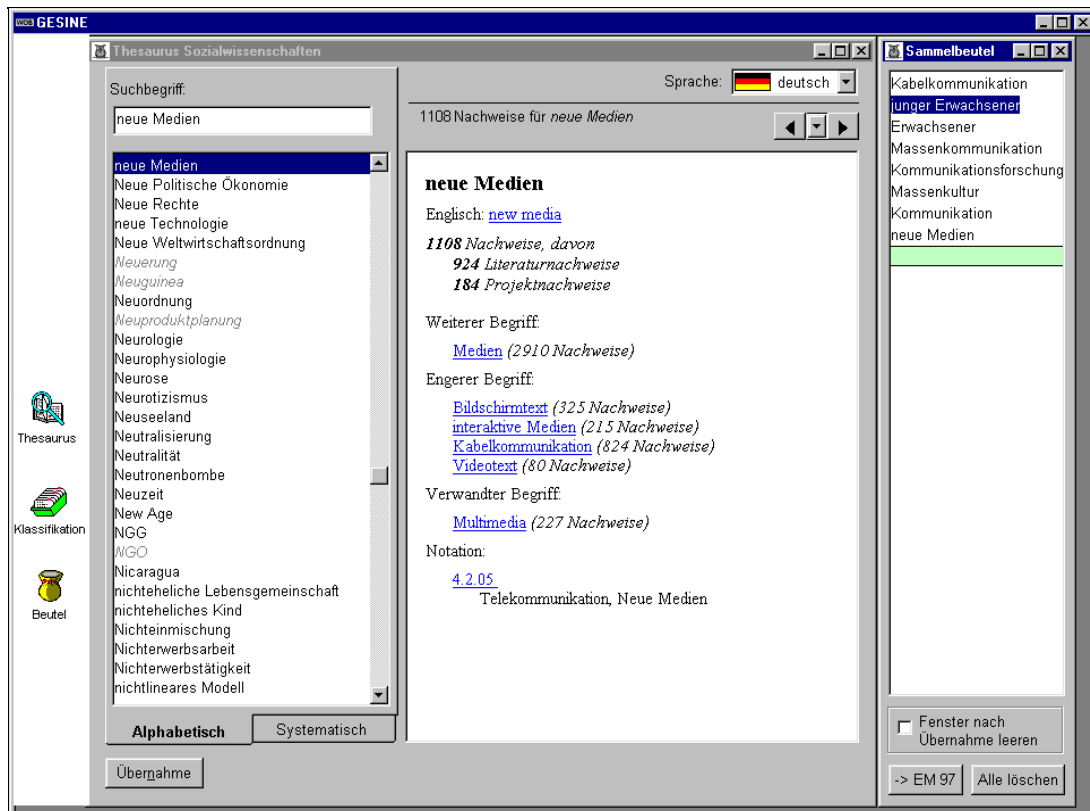


Abbildung 3: Thesaurus und Sammelbeutel mit angezeigtem Deskriptor

4.3.4 Begriffsliste: Tabpage „Systematisch“

In einer dem Explorer bzw. Dateimanager nachempfundenen Baumdarstellung wird die systematische Struktur des Thesaurus abgebildet. Mit Klick auf die Plus- bzw. Minus-Kästchen vor dem Eintrag in einer Zeile wird der Baum an dieser Stelle auf- bzw. zusammengeklappt. Zu jeder Systematikgruppe erscheinen in alphabetischer Reihenfolge alle Deskriptoren und Nichtdeskriptoren. Eine farbliche Unterscheidung ist in der Baumdarstellung nicht realisiert. Beim Wechsel der Tabpage wird der aktuelle Begriff in seiner alphabetischen Einordnung dargestellt. Das Anzeigefeld ändert sich nicht. Eine inkrementelle Suche im Eingabefeld nach Notationen ist nicht möglich.

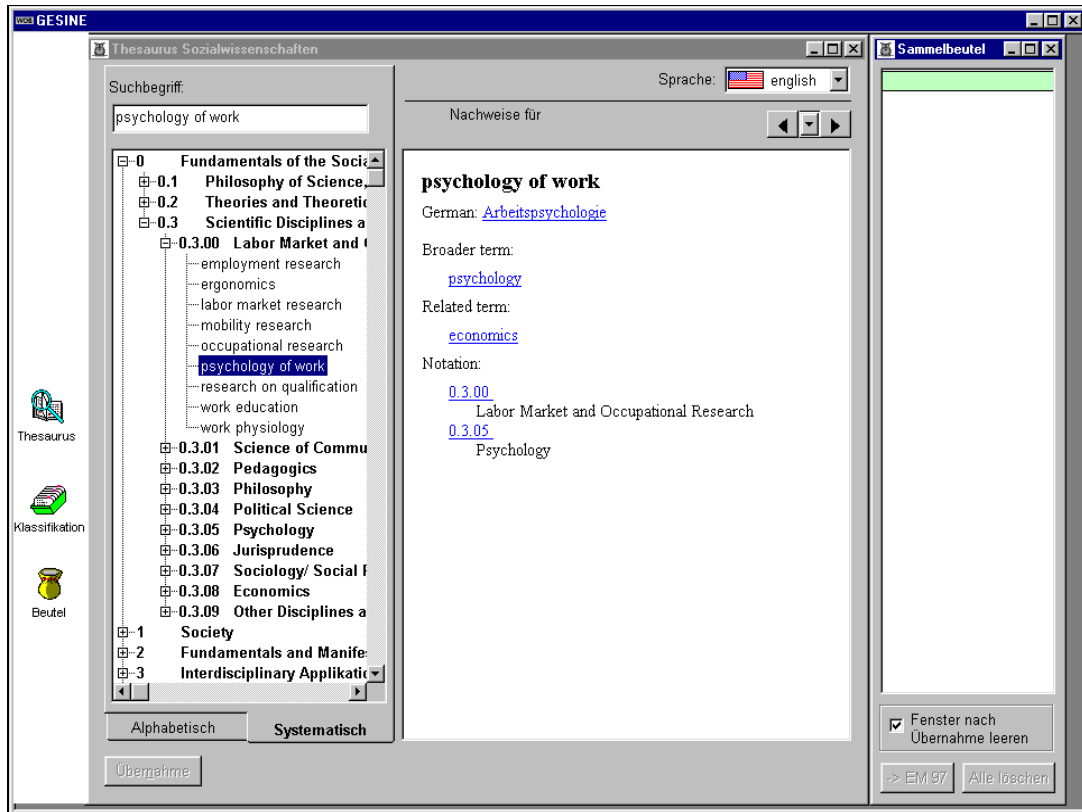


Abbildung 4: Systematische Darstellung eines Nichtdeskriptors

4.3.5 Anzeigefeld

Im Anzeigefeld werden alle zu einem Begriff vorhandenen Informationen angezeigt:

- gültig ab ...
- Übersetzung in die jeweils andere Sprache
- Erläuterung / scope note
- Anzahl Nachweise in den Datenbanken SOLIS und FORIS sowie Gesamtnachweise
- engere/narrower, weitere/broader sowie verwandte/related Begriffe/terms mit Anzahl der Nachweise in SOLIS und FORIS insgesamt

- use ... / benutzte stattdessen ... - Verweise von Nichtdeskriptoren auf Deskriptoren, gekennzeichnet mit rotem Ausrufezeichen
- used for ... /steht für - Verweise von einem Deskriptor auf einen Nichtdeskriptor
- Kombinationsbegriffe (angezeigter Deskriptor steht in Kombination mit weiteren Deskriptoren für einen anderen Begriff)
- Notation

Die Anzeige ist als Hypertext mit zahlreichen Links aufgebaut. Blau dargestellte Texte sind klickbar, der Thesaurus wechselt seine gesamte Anzeige (auch die Sprache) in Abhängigkeit vom geklickten Begriff. Ein einfaches Navigieren in der Thesaurusstruktur ist somit möglich. Beim Klick auf eine Notation wird zur Tabpage „Systematisch“ gewechselt.

4.3.6 Nachweisanzeige

Oberhalb des Anzeigefeldes wird die jeweils aktuelle Trefferzahl für die Zeichenfolge des Eingabefeldes angezeigt. Falls diese kein Begriff der Thesaurus-Liste ist, wird standardmäßig rechts trunziert und die dazugehörige Trefferzahl berechnet. Die Trefferzahlen sind mit Stand vom September 1997 den Produktionsdatenbanken entnommen.

4.3.7 Blätter-Buttons

Rechts oberhalb des Anzeigefeldes ist eine Protokollfunktion für die aktuelle Thesaurussitzung implementiert. Die im Verlauf der Sitzung „besuchten“ Begriffe werden in chronologischer Reihenfolge in einer Liste abgelegt. Diese kann mit dem Pfeil-Button aufgeklappt werden. Neben dem direkten Klick auf den Eintrag kann mit den Vorwärts- und Rückwärts-Blätter-Buttons der Vorgänger bzw. Nachfolger zum aktuell angezeigten Begriff angesprungen werden. Ein „besuchter“ Begriff wird nur einmal in die Liste übernommen.

4.3.8 Button zur Übernahme

Bei Klick auf den Übernahme-Button wird der aktuell angezeigte Deskriptor in den Sammelbeutel kopiert. Übernommen werden nur Deskriptoren, die noch nicht im Sammelbeutel eingetragen sind. Die Übernahme ist nur für

Deskriptoren in deutsch möglich. Bei der Darstellung des Thesaurus in anderen Sprachen wird der Übernahme-Button deaktiviert.

4.4 Sammelbeutel

Im Fenster „Sammelbeutel“ werden in der realisierten Variante des Exports Deskriptoren für die Übernahme in die Terminalemulation EMF[®]97-WIN von aDIS zwischengelagert. Bedingungen für eine korrekte Übernahme sind:

- die Terminalemulation muß gestartet sein
- die Verarbeitungsart „U“ (update) in der Variante „I“ (inhaltliche Erschließung) muß gewählt worden sein
- ein SOLIS- oder FORIS-Dokument muß geladen sein
- eine Eingabemaske muß aktiv sein, d.h. nach erfolgter Plausibilitätsprüfung (etwa nach Einsatz der Textanalyse) muß die Maske „Plausibilitätsprüfung“ mit „K“ (Korrektur) wieder verlassen worden sein
- die zulässige Gesamtzahl von 28 Deskriptoren pro Dokument darf nicht überschritten werden

Die Erfüllung dieser Bedingungen wird vom Programm überprüft, Fehler werden durch Messageboxes angezeigt.

Die Übernahme der Deskriptoren vom „Sammelbeutel“ in die Terminalemulation ist aufgrund der nicht konstanten Reaktionszeiten und der Maskenstruktur von aDIS sowie der Art des Datenaustausches zwischen den beiden Programmen langsamer als Kopiervorgänge zwischen anderen Programmen. Bei der Übernahme werden Umlaute in den Deskriptoren aufgelöst, die die aDIS-Systemprüfung so verarbeitet.

Nach erfolgter Übernahme bleibt das Fenster der aDIS-Anwendung zur weiteren Bearbeitung im Vordergrund. Der Thesaurus wird in den Ausgangszustand zurückgesetzt (Cursor im leeren Eingabefeld), das Protokoll der aktuellen Sitzung (siehe Blätter-Buttons) bleibt erhalten. In windowsüblicher Weise (Tastenkombination <ALT>+<TAB> oder über die Task-Leiste bei Windows 95) kann wieder zum Thesaurus gewechselt werden.

4.4.1 Listenfeld für Deskriptoren

Nach Betätigung des Übernahme-Buttons im Thesaurus wird der dort aktuelle Deskriptor in das Listenfeld des Sammelbeutels eingetragen. Vorab erfolgt eine Dublettenprüfung, so daß für eine Füllung der Liste jeder Deskriptor nur einmal via Button übernommen werden kann. Alle übernommenen Deskriptoren können bei Bedarf editiert werden.

Im Listenfeld des Sammelbeutels befindet sich an jeweils letzter Position eine grün unterlegte Zeile. Diese ermöglicht die direkte Eingabe von Begriffen in den Sammelbeutel. Das ist bei kurzen Deskriptoren (bspw. BRD, ÖTV) schneller als die Übernahme nach Suche in der Begriffsliste. Eine Prüfung auf korrekte Schreibweise findet bei der direkten Eingabe erst mit der aDIS-Systemprüfung statt.

4.4.2 Fenster nach Übernahme leeren

Falls nach Übernahme der eingetragenen Deskriptoren das Listenfeld im Sammelbeutel automatisch geleert werden soll, kann das über die Checkbox im Sammelbeutel eingestellt werden.

4.4.3 Button zur Übernahme in die Terminalemulation

Mit Klick auf den Button ->em97 wird die Übernahme aller Einträge der Liste in die aDIS-Anwendung ausgelöst. Die Deskriptoren werden der Reihe nach in die freien Felder der Maske „Deskriptoren“ eingefügt, für überlange Begriffe wird die Felderweiterung benutzt.

4.4.4 Löschen-Button

Alle Einträge der Liste werden gelöscht.

4.5 Klassifikation

Der Browser zur Darstellung der Klassifikation Sozialwissenschaften ist in analoger Weise zum Thesaurusbrowser konzipiert und aufgebaut.

Das *Eingabefeld* ermöglicht

- bei Einstellung der Tabpage „Begriffe“ die inkrementelle Suche im alphabetisch sortierten Index. Dieser enthält die Überschriften der Klassen sowie Bezeichnungen, die auf Klassifikationen verweisen. Solche Verweise sind kursiv und grau dargestellt.
- bei Einstellung der Tabpage „Hierarchie“ die Suche nach einem Klassifikationscode. Der systematische Aufbau der Klassifikation wird durch eine Baumstruktur verdeutlicht.

Eine *Protokolliste* mit Blättermöglichkeit ermöglicht die Rekonstruktion eines Suchvorgangs in seiner chronologischen Abfolge. Die Liste befindet sich oberhalb des Anzeigefeldes. Der *Übernahme-Button* ist derzeit funktionslos.

Die *Anzeige* ist als Hypertext mit entsprechenden blau dargestellten Links strukturiert. Hier wird die aktuelle Klassifikation in ihrer hierarchischen Einordnung abgebildet. Die Anzahl der dazu vorhandenen Nachweise in SOLIS und FORIS, getrennt nach Vergabe als Haupt- und Nebenklassifikation, wird berechnet. Ferner werden alle Begriffe aufgeführt, für die diese Klassifikation verwendet wird.

Beim Aufruf eines Begriffs, der auf eine Klassifikation verweist, wird dieser oberhalb einer Geraden grau dargestellt. In die Protokolliste wird die dazugehörige Klassifikation festgehalten.

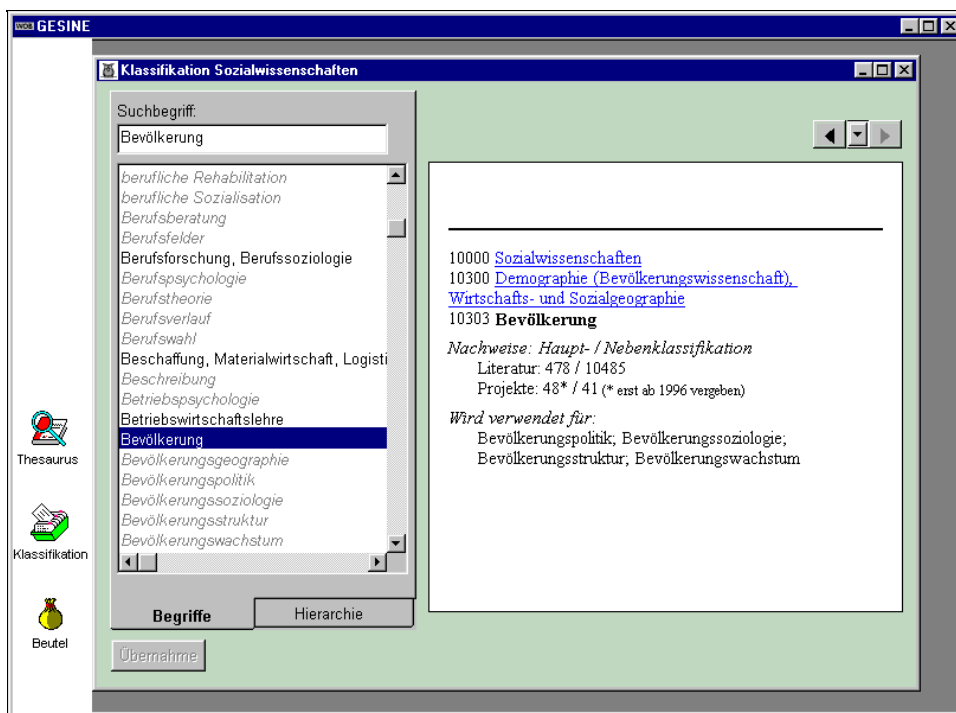


Abbildung 5: Klassifikation, Indexliste

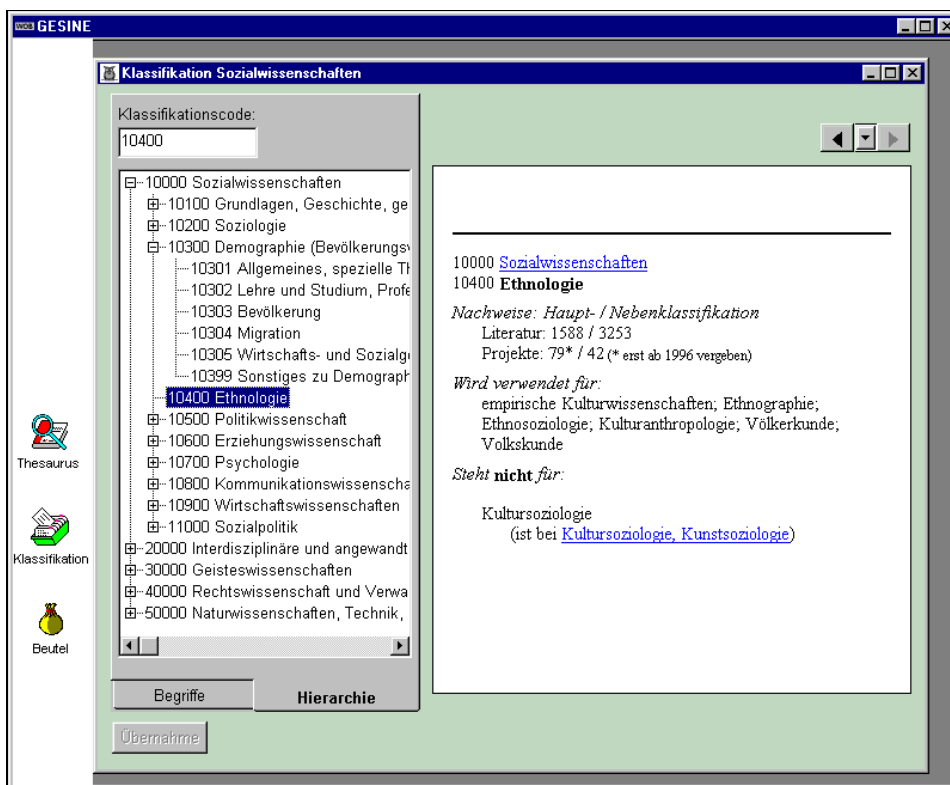


Abbildung 6: Klassifikation, hierarchische Darstellung

4.6 Fenstergröße und -position

Alle Fenster können in der in Windows üblichen Weise mit der Maus in Größe und Position verändert werden. Bestimmte vorgegebene Minimalgrößen können dabei nicht unterschritten werden. Nach Schließen und erneutem Öffnen nehmen alle Fenster ihre voreingestellte Größe und Position ein. Eine Speicherung der individuellen Oberflächen-Konfiguration ist derzeit nicht möglich.

5 Fazit und Ausblick

Nach fast einjährigem Einsatz in der täglichen Arbeit des Informationszentrums Sozialwissenschaften kann man eine positive Bilanz ziehen. Der Thesaurus-Browser mit integrierter Klassifikation und Exportfunktion nach aDIS hat sich als eigenständiger Teil des Programmpakets GESINE bewährt. Durch die Nutzung des Programms kam es für die Anwender zu einer Arbeitserleichterung und Vereinfachung bisheriger Abläufe, was sich nachhaltig bei der Akzeptanz widerspiegelte. Die Vorteile überwogen eindeutig im Vergleich zu den Nachteilen, wie vermehrte Bildschirmarbeit.

Trotz des sehr erfolgreichen Einsatzes und der hohen Akzeptanz sind eine Reihe von Aufgaben für die Programmverbesserung noch offen. Diese werden im Rahmen der Produktbetreuung im Institut zu realisieren sein. Im Einzelnen betrifft das die nachfolgend aufgeführten Unzulänglichkeiten.

Thesaurus

- Es sollten weitere Programmoptimierungen versucht werden, um die Anzeigegeschwindigkeit zu erhöhen. Der jetzige Stand (ca. 0,4 Sekunden pro Begriff, siehe Tabelle 1) ist zwar akzeptabel, aber noch verbesserungsbedürftig.
- Das Problem der korrekten Aktualisierung der Anzeigen bei Begriffen mit Umlauten, welches im Algorithmus der inkrementellen Suche liegt, muß gelöst werden. Andere Anbieter umgehen dies, indem Umlaute grundsätzlich aufgelöst werden.

- Das verwendete control zur Anzeige der Begriffsrelationen vermag lediglich eine auf HTML 2 basierende Anzeige zu realisieren. Eine Umstellung auf HTML 3 ist wegen besserer Formatierungsmöglichkeiten vorgesehen.
- Das verwendete control für die Baumdarstellung ist in seiner Anzeige auf eine für die Belange des Thesaurus nicht ausreichende Anzahl von Einträgen beschränkt. Eine verbesserte Version wird nach Erscheinen implementiert.
- Auf Wunsch der Anwender der Abteilung Informationstransfer Osteuropa des IZ sollte der Thesaurus auch in russischer Sprache verfügbar sein. Dazu sind die Daten maschinenlesbar vorhanden. Die Möglichkeiten zur Implementation kyrillischer Zeichensätze in PowerBuilder-Applikationen müssen dazu geprüft werden, ebenso die Umschaltung Anzeige in den verwendeten controls.

Klassifikation

- Der für die Klassifikation benutzte Index enthält in den verwendeten Überschriften Texte, die durch Auflösung in Einzelbegriffe an Informationskraft gewinnen könnten. Damit würden sich zugleich die Suchmöglichkeiten im Index erweitern. Solche Klassifikationsgruppen wie „Allgemeines zur ...“, „Ethik der ...“, „Geschichte der ...“, „spezielle Theorien der ...“, „Studium der ...“ oder „Sonstiges zur ...“ sollten intellektuell aufgelöst werden. Ebenso könnten zusätzliche Indexbegriffe mit Verweisen auf zu verwendende Klassifikationen ergänzend hinzugefügt werden.
- Auch hier ist das verwendete control für die Baumdarstellung durch eine verbesserte Version zu ersetzen.

Sammelbeutel

- Die Exportfunktionalität, die zunächst für die institutsinterne Verwendung entwickelt wurde, ist zu verallgemeinern. Damit wäre der Sammelbeutel auch in anderen Anwendungsfällen verwendbar.
- Durch Kapselung der Funktionalität zum Datenaustausch mit aDIS könnten auch im Institut weitere Anwendungsfälle bedient werden. Bspw. könnte so die Dublettenprüfung in aDIS im Zusammenhang mit der Literaturbeschaffung erleichtert werden. Aus den maschinenlesbaren Neuerscheinungslisten kann durch Übergabe des Titelstrings oder der ISBN-

Nummer in aDIS ermittelt werden, ob die betreffende Literatur bereits zur Auswertung bestellt wurde.

- Eventuell sind mit dem Einsatz einer neuen aDIS-Version Anpassungsarbeiten nötig.

6 Literatur

Frisch, E.; Kluck, M. (1997): Pretest zum Projekt German Indexing and Retrieval Testdatabase (GIRT) unter Anwendung der Retrievalsysteme Messenger und freeWAISsf. IZ-Arbeitsbericht Nr. 10, Informationszentrum Sozialwissenschaften Bonn

Krause, J. (1995): Das WOB-Modell. IZ-Arbeitsbericht Nr. 1, Informationszentrum Sozialwissenschaften Bonn

Krause, J.; Zimmer, M. (Hrsg) (1997): Informationsservice des IZ Sozialwissenschaften. Datenbankentwicklung und Nutzung, Netzwerke, Wissenschaftsforschung. Informationszentrum Sozialwissenschaften Bonn

Marx, J.; Mutschke, P.; Schommler, M. (1995): Möglichkeiten der intelligenten Integration heterogener Datenbestände: das Projekt GESINE. IZ-Arbeitsbericht Nr. 2, Informationszentrum Sozialwissenschaften Bonn

Mutschke, P. (1996): Documentation of Conceptual and Physical Data Modell IZ-Datenbank. IZ-Arbeitsmaterial Nr. 5, Informationszentrum Sozialwissenschaften Bonn

Riege, U.; Schomisch, S.; Schommler, M. (1996): COGET - Computergestützte Erstellung von Themendokumentationen. IZ-Arbeitsbericht Nr. 9, Informationszentrum Sozialwissenschaften Bonn

Thesaurus Sozialwissenschaften /Alphabetischer Teil, Informationszentrum Sozialwissenschaften Bonn 1996

Thesaurus Sozialwissenschaften /Systematischer Teil, Informationszentrum Sozialwissenschaften Bonn 1996

Thesaurus for the social sciences: Englisch-German, Informationszentrum Sozialwissenschaften Bonn 1997

Thesaurus Sozialwissenschaften: Deutsch-Englisch, Informationszentrum Sozialwissenschaften Bonn 1996

Thesaurus Sozialwissenschaften; Band 1: Deutsch-Englisch-Russisch; Band 2: Russisch-Deutsch-Englisch; Band 3: Register; Bonn, Moskau 1997

Geschäftsstelle des Wissenschaftsrates (1997): Wissenschaftspolitische Stellungnahme zur GESIS - Gesellschaft Sozialwissenschaftlicher Infrastruktureinrichtungen e.V.. Berlin 1997