

CARMEN

Content Analysis, Retrieval and MetaData: Effective Networking

Abschlussbericht des Arbeitspakets 12 (AP 12)
Crosskonkordanzen von Klassifikationen und Thesauri

Als Online-Publikation aufbereitete Version

Zuwendungsempfänger:

Universität Regensburg

Informationszentrum Sozialwissenschaften (s.a. AP 11)

Die Deutsche Bibliothek (s.a.AP 4)

Deutsches Institut für Internationale Pädagogische Forschung

Vorhabensbezeichnung:

Crosskonkordanzen von Klassifikationen und Thesauri (CARMEN-AP 12)

Laufzeit des Vorhabens:

01.12.1999 - 31.03.2002

Berichtszeitraum:

01.12.1999 - 31.03.2002

1. Aufzählung der wichtigsten wissenschaftlich-technischen Ergebnisse und anderer wesentlicher Ereignisse

Zusammenfassung

Ziel des Arbeitspakets 12 (AP 12)¹ innerhalb der Sonderfördermaßnahme CARMEN war eine integrierte Suche nach sachlichen Gesichtspunkten in verteilten Datenbeständen mit unterschiedlichen inhaltlichen Schwerpunkten zu ermöglichen. Dabei soll den begrifflichen Unterschieden der verwendeten Thesauri und Klassifikationen durch Crosskonkordanzen Rechnung getragen werden.

Zur Erreichung des Gesamtziels war erforderlich

- die Untersuchung der Methodik von Crosskonkordanzen zwischen Klassifikationen bzw. Thesauri.
- die Programmierung eines Verfahrens, wie solche Crosskonkordanzen zwischen unterschiedlichen im Internet verfügbaren Klassifikationen bzw. Thesauri abgebildet werden können.
- die prototypische Erarbeitung solcher Crosskonkordanzen für bestimmte Fachgebiete und ausgewählte Klassifikationen bzw. Thesauri.

Einordnung in das Gesamtkonzept von CARMEN

Ausgangspunkt ist die Vergabe von Metadaten durch Autoren bzw. Bibliotheken und Fachinformationszentren nach Dublin Core. Diese basiert auf bereits entwickelten Klassifikationssystemen bzw. Thesauri oder bei Daten von geringerer Erschließungsintensität auf frei vergebenen Deskriptoren. Notwendig ist eine durchgängige, interoperable Nutzung von Metadaten. Durch die Erarbeitung von Crosskonkordanzen wird die Nutzung von Metadaten bei elektronischen Publikationen gefördert, ebenso wie die Recherche in online verfügbaren Bibliothekskatalogen und Fachdatenbanken.

Im Gesamtrahmen von CARMEN realisierte AP 12 eines der Transfermodule von AP 11, wo die zu integrierenden Textgrundlagen in bezug auf die verbleibende Heterogenität nach dem Wirksamwerden der postulierten DC-Maßnahmen analysiert und entsprechende Algorithmen zu ihrer Behandlung realisiert wurden. Crosskonkordanzen unterstützen den Ausschnitt des Transfers, der sich auf die generalisierbaren Relationen des Wortschatzes bezieht. AP 12 befasste sich also damit, intellektuell erschlossene Datenbestände gemeinsam suchbar zu machen, während AP 11 sich mit maschinellen Verfahren befasste.

Im Hinblick auf die spätere Nachanwendung waren zwei Ziele gestellt:

- die prototypische Erarbeitung eines Gesamtsystems zur Verbesserung des Retrievals nach elektronischen Publikationen. AP 12 war insofern AP11 zugeordnet, in dem zusammen mit AP 7 das lauffähige Gesamtsystem erarbeitet wurde.
- Andererseits verfolgte AP 12 das Ziel, selbst lauffähige Module zu erarbeiten, die für andere Retrievalsysteme genutzt werden können.

Beide Ziele wurden voll erreicht.

¹ Homepage zum Projekt: <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html>

Das Szenario des Einsatzes von AP 12 umfasst

- die verteilte Suche in verschiedenen Bibliothekskatalogen, für die in der Zwischenzeit verschiedene kommerzielle und von Institutionen entwickelte Software besteht (KVK, IPS von IHS, Elektra von Sisis, Metalib von Exlibris). Allerdings ist in allen Systemen bisher keine sachliche Suche realisiert, da die entsprechenden Transfermodule fehlen,
- den Übergang in der Suche zwischen Fachdatenbanken und allgemeinen Bibliothekskatalogen,
- die kombinierte Suche in mehreren fachlich unterschiedlichen Fachdatenbanken oder virtuellen Bibliotheken.

Im Arbeitspaket wurden prototypisch für alle drei Szenarien Crosskonkordanzen erarbeitet (vgl. 1.3 und 1.4).

Die Klassifikationen und Thesauri wurden nach dem Umfang ihrer Anwendung und der zugänglichen Daten ausgewählt. Es handelt sich jeweils um Klassifikationen und Thesauri von hoher Relevanz.

RVK (Regensburger Verbund-Klassifikation) ist die in deutschen Bibliotheken bei weitem verbreiteste Klassifikation (über 50 Anwender in Hochschulbibliotheken). Sie ist ursprünglich für die Aufstellung von Buchbeständen entstanden, wird aber heute in gleichem Umfang für die Recherche in Datenbanken verwendet.

BK (Basisklassifikation) wird in GBV und in einzelnen Bibliotheken außerhalb verwendet.

MSC (Mathematical Subject Classification) wird weltweit und sehr verbreitet in Erschließungssystemen für die Mathematik verwendet.

PACS (Physics and Astronomy Classification Scheme) wird weltweit in Erschließungssystemen für die Physik verwendet. Dadurch dass es verschiedene Versionen gibt, ist die Anwendung jedoch eingeschränkter als bei MSC.

IZK (IZ-Klassifikation) ist die Fachklassifikation des IZ für die Recherche in Online-Datenbanken. Es handelt sich um die größten deutschen Fachdatenbanken der Sozialwissenschaften.

SWD (Schlagwortnormdatei) ist Standard für die verbale Indexierung der deutschen Bibliotheken. Der Thesaurus wird kooperativ geführt.

TheSoz (Thesaurus Sozialwissenschaften) ist der Thesaurus des IZ, d. h. der größten deutschen Fachdatenbanken.

ThesBild (Schlagwortmaterial der FIS Bildung). Hier handelt es sich um den größten deutschen Thesaurus der Erziehungswissenschaften.

DDC (Dewey Decimal Classification) ist die international am weitesten verbreitete Klassifikation. Sie konnte aus urheberrechtlichen Gründen nicht einbezogen werden. Daher wurde sie im Laufe des Projekts durch die BK ersetzt.

Anzumerken ist, dass Ziel des Arbeitspakets nicht eine Überarbeitung der bestehenden Klassifikationen bzw. Thesauri war, obwohl dies sicherlich ein dankbares Verfahren zur Überwindung der Heterogenität der Erschließung wäre. Zu beobachten war z. B. dass während der Laufzeit des Projekts die DDC an MSC angepasst wurde. In der Regel wird man aber eine solche Überarbeitung nicht anstreben können. Methodisch geht es um die Suche in heterogenen Sacherschließungssystemen.

Die Ziele des Projekts wurden mit einem lauffähigen Gesamtsystem und mit weiter nutzbaren Crosskonkordanzen erreicht.

Die Retrievaltests ergaben, dass insgesamt sinnvolle Ergebnisse mit den durchgeführten Verfahren erzielt werden können. Die Relevanz der Treffer ist naturgemäß unterschiedlich je nach der verwendeten Klassifikation bzw. Thesaurus. Generell kann gesagt werden, dass die

Qualität um so höher ist, je eher die Strukturprinzipien der Klassifikationen bzw. Thesauri vergleichbar sind. Dadurch dass teilweise in Randbereichen von Fachthesauri/Fachklassifikationen gröber erschlossen wird als in Kernbereichen und durch den Übergang beim erschlossenen Material (Bücher einerseits, Zeitschriftenaufsätze andererseits) ergeben sich Einschränkungen. Trotzdem hat sich das Verfahren bewährt und sollte breit angeboten werden.

Für Crosswalks zwischen Thesauri und Klassifikationen wurden erste Untersuchungen vorgenommen. Diese Ergebnisse lassen es auch sinnvoll erscheinen, auf diesen Weg weiter voranzugehen.

Die bei der Erarbeitung der Konkordanzen verwendeten Relationen haben sich insgesamt bewährt.

Update der Crosskonkordanzen

Bei Klassifikationen stellt dies insofern kein umfangreiches Problem dar, als Klassifikationen in der Regel in größeren Zeiträumen überarbeitet werden (z. B. MSC alle 10 Jahre, DDC alle 4 Jahre, RVK kontinuierlich jedoch in begrenztem Umfang von ca. 200 Systemstellen jährlich). Das Konkordanzmodul erlaubt technisch ein Update. Das Klassifikationsmodul kann unterschiedliche Versionen noch nicht verwalten.

Größere Probleme entstehen bei Thesauri, wenn sie für Änderungen offen sind. Dies gilt insbesondere für die SWD, die kontinuierlich stark erweitert und um Informationen angereichert wird. Hierfür ist mit SIS-TMS noch keine optimale Lösung gefunden.

Nutzung und Nachnutzung

Die im Projekt erarbeiteten Konkordanzen im Bereich Klassifikation werden im Web bereitgestellt. Das verwendete Konkordanzmodul CarmenX wurde im EU-Projekt Renardus sowie in einer ganzen Reihe weiterer Projekte nachgenutzt. Es ist beabsichtigt das Tool weiter zu entwickeln und für virtuelle Bibliotheken einzusetzen.

Die Crosskonkordanzen zwischen Thesauri sind in die Virtuelle Bibliothek Sozialwissenschaften eingebunden und sollen in den Informationsverbund Sozialwissenschaften-Bildungspsychologie integriert werden.

Vorbemerkung zur Gliederung

Im Antrag wurden zum einen UAP-Nr. nach Partnern differenziert vergeben (d.h. auch verschiedene Nummern bei Kooperationen zu einer inhaltlichen Aufgabe), und zum anderen wurden inhaltlich definierte Arbeitsphasen vorgegeben. Der vorliegende Abschlussbericht orientiert sich zum besseren Verständnis an den inhaltlich definierten Arbeitsphasen und ordnet die jeweils betroffenen UAP-Nr. zu.

1.1. Konzeptuelle und inhaltliche Grundlagen (UAP 1-4, 6-8, 11)

Die Definition der Aufgaben, die Synopse der Datenformate und die Datenbereitstellung wurden auf den Arbeitsbesprechungen vom 01.02.2000 und 09.03.2000 behandelt und eine Absprache über die konkrete Aufgabenverteilung getroffen.

Für die Erarbeitung der Crosskonkordanzen wurden im Projektverlauf folgende Klassifikationen und Thesauri herangezogen:

a) Klassifikationen

- Basisklassifikation (BK bzw. GBV/PICA)
- Klassifikation Sozialwissenschaften des IZ (IZK)
- Mathematics Subject Classification (MSC)
- Physics and Astronomy Classification Scheme (PACS)
- Regensburger Verbundklassifikation (RVK)
- Dewey Decimal Classification (DDC)

Zur DDC wurden, wie weiter unten erläutert, keine Crosskonkordanzen erstellt.

b) Thesauri

- Schlagwortnormdatei (SWD)
- Thesaurus Sozialwissenschaften des IZ (TheSoz)
- Schlagwortmaterial des FIS Bildung (ThesBild)

Eine parallele Vorgehensweise bei Klassifikationen und Thesauri ermöglicht es, Erkenntnisse über die gemeinsamen bzw. abweichenden Probleme und Lösungsmethoden verschiedener Erschließungsverfahren und Fachgebiete zu gewinnen. Dies stellt den prototypischen Charakter der Untersuchung sicher. Die Lösungen sollen auch auf andere, nicht in die Untersuchung einbezogene Klassifikationen und Thesauri übertragen werden können. Gemeinsam für beide Unterbereiche Klassifikation und Thesaurus ist die Klärung methodischer Fragen.

Die Crosskonkordanzen verweisen auf Klassifikationen/Thesauri, die in sich geschlossene Systeme darstellen. Zwischen verschiedenen Klassifikationen/Thesauri muss mit Hilfe der Crosskonkordanz navigiert werden können. Grundlegende Tatsache ist, dass die Klassifikationen/Thesauri von unterschiedlichen Einrichtungen geführt werden und nicht von der Einrichtung, welche die Konkordanz betreibt. Ein Zugriff muss also auch über das Web auf fremde Rechner möglich sein. Dies vereinfacht auch die Probleme des Updates, ein wichtiges Problem, wenn man den dauerhaften Betrieb eines solchen Systems betrachtet. Allerdings setzt dies voraus, dass die Systeme im Web verfügbar sind (was für die DDC nur sehr bedingt gilt) und entsprechend aufbereitet sind.

Wesentlich für die Datenstruktur ist, dass die Konkordanz stets zwischen zwei Klassifikationen/Thesauri gebildet wird und nicht alle Klassifikationen/Thesauri auf eine/einen der verwendeten oder eine neue, eine Metaklassifikation/einen Metathesaurus abgebildet werden. Die Verknüpfung erfolgt bei den Thesauri über die Deskriptoren, bei den Klassifikationen über die Notationen. Die Verknüpfung hat dabei nach unseren Vorstellungen eine Richtung,

d.h. es ist denkbar, dass die Verknüpfung in der Gegenrichtung anders angelegt wird. Die Konkordanzen werden in der Regel von zwei Richtungen ausgehend bearbeitet. Die Konkordanz zwischen SWD und dem Schlagwortmaterial des FIS Bildung wird nur in der Richtung ThesBild → SWD angelegt. Zusätzliche, über die Thesaurusrelationen hinausgehende Verknüpfungen in den Thesauri (in der SWD beispielsweise Notationen, Ländercode, Zeitangaben) wurden bei der Konkordanzerstellung vernachlässigt.

Bei den Klassifikationen kann die Hierarchie benutzt werden, um die Anzahl der Konkordanzbeziehungen zu verringern. Aus einer Verknüpfung zweier Klassifikationen lassen sich aus den Hierarchiebeziehungen weitere Konkordanzbeziehungen ableiten. Die abgeleiteten Verknüpfungen werden, da redundant, nicht gespeichert.

Neben der Verknüpfung selbst muss die Art der Beziehung zwischen einander zugeordneten Notationen/Deskriptoren erfasst werden:

- Äquivalenz (Symbol: =)
- Oberbegriff (Symbol: >)
- Unterbegriff (Symbol: <)
- Verwandter Begriff (Symbol: ><)

Ein weiterer Punkt, der bei jeder Beziehung erfasst werden muss, ist die Relevanz der Verknüpfung. Diese ist ein grober Schätzwert. Wir unterteilen ihn daher nur in 3 Stufen:

- hoch
- mittel
- gering

Dabei kann man Precision und Recall empirisch ermitteln oder zu schätzen versuchen.

1.2 System zur Datenhaltung und –bearbeitung (UAP 2 , 3, 6, 12)

1.2.1 Softwaremodul zur Bearbeitung von Klassifikationen und Crosskonkordanzen bei Klassifikationen

Die Entscheidung, für das Projekt ein eigenes Werkzeug für die Bearbeitung von Klassifikationen und Crosskonkordanzen zwischen Klassifikationen zu erstellen, basierte auf folgenden Überlegungen:

Ein manuelles Erstellen von Crosskonkordanzen sollte auf jeden Fall vermieden werden. Ein listenorientiertes Arbeiten mit einem Tabellenkalkulationsprogramm, wie bei den Thesauri angewandt, ist wegen der Bedeutung der hierarchischen Gliederung bei Klassifikationen nicht zweckmäßig. Das Werkzeug sollte möglichst schnell zur Verfügung stehen, damit rasch mit der intellektuellen Arbeit begonnen werden konnte. Parallel dazu zeichnete sich in der Anfangsphase des Projektes ab, dass für die Erstellung der Crosskonkordanzen bei Thesauri das Thesaurus-Verwaltungsprogramm SIS-TMS beschafft werden wird.

SIS-TMS wies, da es für Thesauri konzipiert war, für die Bearbeitung von Klassifikationen einige Defizite auf. Als Beispiel zwei der Probleme:

- SIS-TMS setzt voraus, dass die Benennung eindeutig ist. Für Deskriptoren trifft dies zu, nicht jedoch für die Benennung von Systemstellen.
- In Klassifikationen haben Systemstellen eine Reihenfolge, die durch die Sortierreihenfolge der Notationen bestimmt wird. Da Thesauri eine bestimmte Abfolge der Deskriptoren nicht kennen, wird dies von SIS-TMS auch nicht unterstützt.

Aus den genannten Gründen erschien es am zweckmäßigsten, ein Werkzeug basierend auf Vorarbeiten in anderen Bereichen mit gleicher Technologie zu erstellen, um möglichst schnell mit der Arbeit beginnen zu können. Davon ausgehend wurde ein Werkzeug für Crosskonkordanzen von Klassifikationen erstellt, das am 8.3.2000 in einer Testversion freigegeben

werden konnte. Nach einer Testphase und einigen Änderungen konnte am 27.03.2000 mit der Arbeit an den Crosskonkordanzen begonnen werden.

Parallel zur intellektuellen Arbeit wurde die Software im Laufe des Projekts konsolidiert und verbessert. Durch die Erfahrungen mit anderen Klassifikationen konnte das Modell für die Klassifikationen im Programm verallgemeinert werden. Als Ergebnis steht jetzt ein Werkzeug zur Verfügung, das ohne größere Anpassungen für eine Reihe von Klassifikationen eingesetzt werden kann. Spezifische Merkmale der Klassifikationen werden in Treiber-Dateien ausgelagert, so dass neue Klassifikationen leicht eingebunden werden können.

Sowohl das Modul zur Erstellung von Crosskonkordanzen, als auch das Modul zur Anzeige einer Klassifikation basieren auf derselben Software. Die Abstraktion der Software auf eine verallgemeinerte Klassifikation gelang ohne Abstriche, so dass die im Projekt programmierte Software mittlerweile auch für die Anzeige und Bearbeitung der Regensburger

Verbundklassifikation eingesetzt wird (http://www.bibliothek.uni-regensburg.de/rvko_neu/).

Genauso kann das Programm auch für alle anderen in diesem Projekt bearbeiteten Klassifikationen genutzt werden und darüber hinaus für eine Reihe weiterer Klassifikationen. Für die Software zur Erstellung und Bearbeitung der Crosskonkordanzen wurde der Name CarmenX geprägt.

CarmenX zeichnet sich durch folgende Merkmale aus:

- das Programm ist WWW-basiert
- für die kooperative Bearbeitung durch mehrere Anwender geeignet
- auf Client-Seite ist nur ein WWW-Browser erforderlich
- Konkordanzbeziehungen können unsymmetrisch, d.h. nur einseitig, definiert werden
- die Klassifikationen müssen nicht ins System geladen werden, sondern können auch verteilt im Internet liegen
- für die Nachnutzer fallen keine Lizenzkosten für die Software an

Die Möglichkeit, Konkordanzen einseitig zu bilden, dient der Erstellung von Crosskonkordanzen von zwei verschiedenen Seiten aus. Zweckmäßigerweise arbeitet man die Notationen einer Ausgangsklassifikation systematisch durch und bildet diese auf die Zielklassifikation ab. In einem zweiten Durchgang vertauscht man Ziel- und Ausgangsklassifikation. Wenn idealerweise verschiedene Bearbeiter an einer Konkordanz arbeiten, dokumentieren unsymmetrische Beziehungen unterschiedliche Bewertungen, die später bereinigt werden können. Beim Erstellen der Konkordanzen müssen die Bearbeiter in Zweifelsfällen die Klassifikationspraxis durch Recherchen in Katalogen und Fachdatenbanken studieren. Die Software unterstützt dieses Vorgehen. Mit einem Knopfdruck startet eine Recherche mit der aktuellen Notation als Suchwert im Katalog des Bibliotheksverbunds Bayern (BVB), des Südwestdeutschen Bibliotheksverbunds (SWB) oder im Gemeinsamen Bibliotheksverbund Göttingen (GBV).

Einbindung der Crosskonkordanzen von Klassifikationen in die Gesamtarchitektur von CARMEN.

Die Crosskonkordanzen werden in der Gesamtarchitektur von CARMEN über das in AP 11 entwickelte Transfermodul nutzbar gemacht. Um den Datenfluss zwischen AP 11 und AP 12 möglichst reibungsfrei zu gestalten, hat man sich innerhalb von AP 12 darauf geeinigt, die Crosskonkordanzen von Thesauri und Klassifikationen nur von einem System aus an das Transfermodul von AP 11 zu binden.

Auf ihre Eignung für diesen Zweck wurden die beiden in AP 12 genutzten Systeme, SIS-TMS und CarmenX, geprüft. Miteinbezogen in die Prüfung wurde dazu das im Projekt MACS entwickelte Werkzeug. Das Programm aus MACS schied aus, weil die erforderlichen Anpassungen an AP 12 im Projektzeitraum nicht möglich gewesen wären.

Zum Test von SIS-TMS wurden die Crosskonkordanzen und zugehörigen Klassifikationen aus CarmenX ausgespeichert und in SIS-TMS aufgenommen. In SIS-TMS müssen zwingend auch die Klassifikationen importiert werden. Die Konkordanzen können nicht unabhängig verwaltet werden. Umgekehrt wurden die Crosskonkordanzen der Thesauri aus SIS-TMS ausgespeichert und nach entsprechender Anpassung in CarmenX eingespielt.

Nach dieser Prüfung ergab sich, dass SIS-TMS und CarmenX für die Verwaltung von Crosskonkordanzen von Thesauri und Klassifikationen geeignet sind. Hinsichtlich der Datenstruktur sind Crosskonkordanzen einfacher als Thesauri und Klassifikationen und stellen daher auch weniger Anforderungen an die Software. Weiter zeigte sich, dass weder SIS-TMS für Klassifikationen, noch CarmenX für Thesauri die optimale Unterstützung bot.

Aus praktischen Erwägungen (räumliche Nähe AP 11/AP 12) wurde der Datenfluss nach AP 11 über SIS-TMS angebunden.

1.2.2 Softwaremodul zur Erstellung von Crosskonkordanzen bei Thesauri (in Kooperation mit AP 11)

Das Ausgangsmaterial für die in AP12 intellektuell zu erstellenden Crosskonkordanzen bildeten drei Thesauri:

- die Schlagwortnormdatei (SWD) der Deutschen Bibliothek mit insgesamt rund 142.000 Sachschlagwörtern
- der Thesaurus Sozialwissenschaften (TheSoz) des Informationszentrums Sozialwissenschaften mit rund 11.000 Einträgen (7.200 zugelassenen Schlagwörtern)
- das Schlagwortmaterial des FIS Bildung (ThesBild) in einem Auszug von rund 900 Sachschlagwörtern aus insgesamt 60.000 Einträgen

Selbst wenn man berücksichtigt, dass im Rahmen des Projektes die Crosskonkordanzen nur prototypisch für fachlich relevante Teilgebiete erarbeitet werden können, handelte es sich insgesamt um rund 45.000 Begriffe, die als Basisbestand eingelesen werden mussten. Hinzu kam, dass alle drei Thesauri zu den Termen Verweisungen und Relationen enthalten - Verweisungen auf Vorzugsbenennungen und Relationen, die das begriffliche Umfeld eines Term anzeigen, d.h. weitere, engere und verwandte Begriffe.

Beide Faktoren, die Menge der Terme gut handhaben und das begriffliche Umfeld bei der intellektuellen Bearbeitung von Crosskonkordanzen im Blick halten zu können, waren zentrale Forderungen an die Software. Weitere Anforderungen waren:

- die parallele Verwaltung mehrerer Thesauri
- geeignete Import- und Exportschnittstellen zum Einlesen der Thesauri bzw. zur Weiterverwendung der Daten im Retrieval
- ein Modul zur Erstellung der Interthesaurusbeziehungen (Art der Beziehung und Gewichtung)
- ein Modul zur später evtl. erforderlichen Thesauruspflege
- ein Preis von max. 10.000 DM

Eine Sichtung der auf dem Markt angebotenen Software² zeigte sehr schnell, dass die meisten Thesaurus-Verwaltungsprogramme, auch wenn sie Multilingualität anbieten, letztlich doch nur einen Thesaurus verwalten können. Die Mehrsprachigkeit wird in diesen Fällen über Synonymbeziehungen innerhalb des Thesaurus abgewickelt. Mehrere Thesauri kann die Software INDEX 4.1 verwalten. Sie setzt allerdings voraus, dass man als Basis zusätzlich Lotus Notes zur Verfügung hat, und liegt preislich gesehen damit über dem veranschlagten Limit.

² Ende 1999/ Anfang 2000 wurde die unter http://www.fbi.fh-koeln.de/labor/Bir/thesauri_new/thessoft.htm vorliegende Sammlung von Thesaurus-Software gesichtet

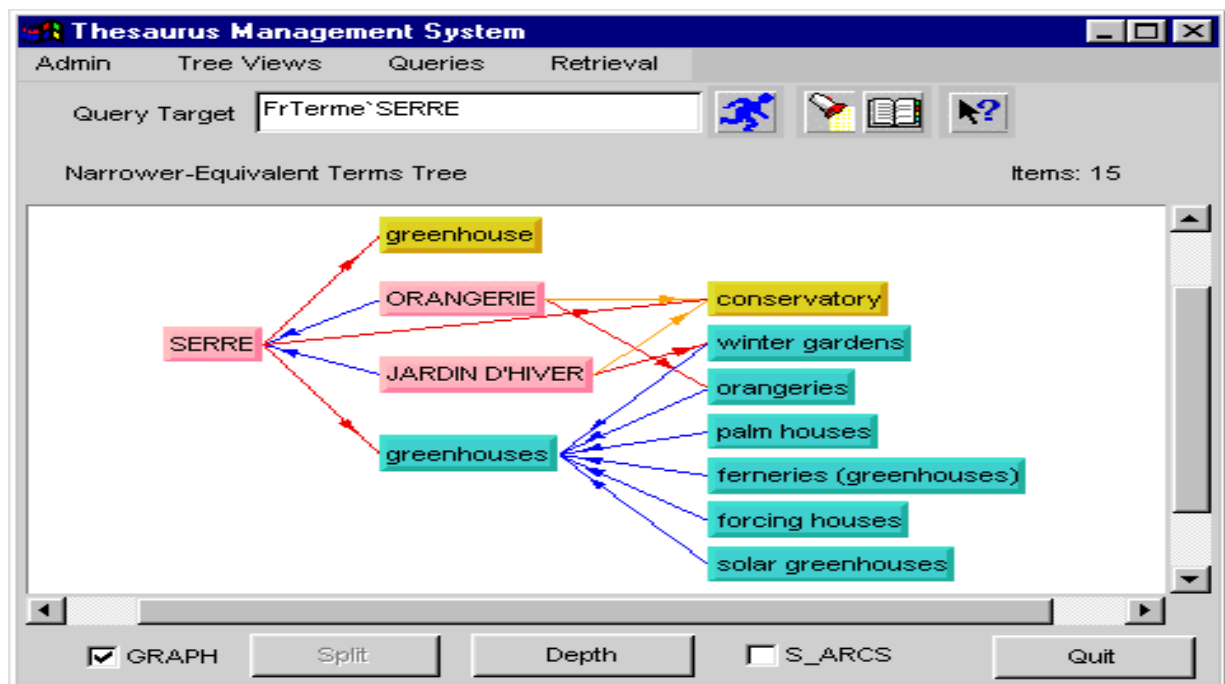
Wir haben uns für SIS-TMS (Thesaurus Management System for Distributed Digital Collections, eine Anwendung des Semantic Index System) entschieden. SIS-TMS wurde vom Centre for Cultural Informatics and Documentation Systems, Institute of Computerscience, Foundation for Research and Technology - Hellas (ICS-FORTH) – entwickelt. Es ist in Englisch oder Französisch erhältlich und kann unter Windows '95, Windows NT, Solaris und AIX laufen. SIS-TMS wird seit einiger Zeit in der Getty Foundation angewendet. Dort werden drei Thesauri mit starken Hierarchien zu Kunst und Architektur (zwei englische und ein französischer) miteinander verzahnt.³

SIS-TMS bietet die Möglichkeit, mehrere Thesauri parallel zu laden, innerhalb eines Thesaurus Deskriptoren und Relationen zu bearbeiten oder Relationen zwischen Begriffen verschiedener Thesauri zu erstellen. Die Thesauri und Interthesaurusbeziehungen werden in einem semantischen Netzwerk gespeichert. Außerdem bietet SIS-TMS ein graphisches User Interface zur Navigation innerhalb und zwischen den Thesauri und zur Präsentation von Relationen.

Beispiele zur grafischen Oberfläche

a) aus der Anwendung in der Getty Foundation:

der französische Term „SERRE“ in der „Term-Position-in-Hierarchy-View“ von SIS-TMS



Die Zugehörigkeit der Begriffe zu den drei Thesauri wird über die Farbe (rot/grün/gelb) deutlich gemacht.

³ Test versions exist with the content of the AAT (Art & Architecture Thesaurus), the thesaurus of the Royal Commission of Historical Monuments of England (RCHME) and of the French MERIMEE database and equivalence relations between the three. Besides others, SIS-TMS has been installed at the French Ministry of Culture, the Greek Ministry of Culture and the Royal Commission of Historical Monuments of England

Farblich gekennzeichnet sind auch die verschiedenen Typen der Relationen:

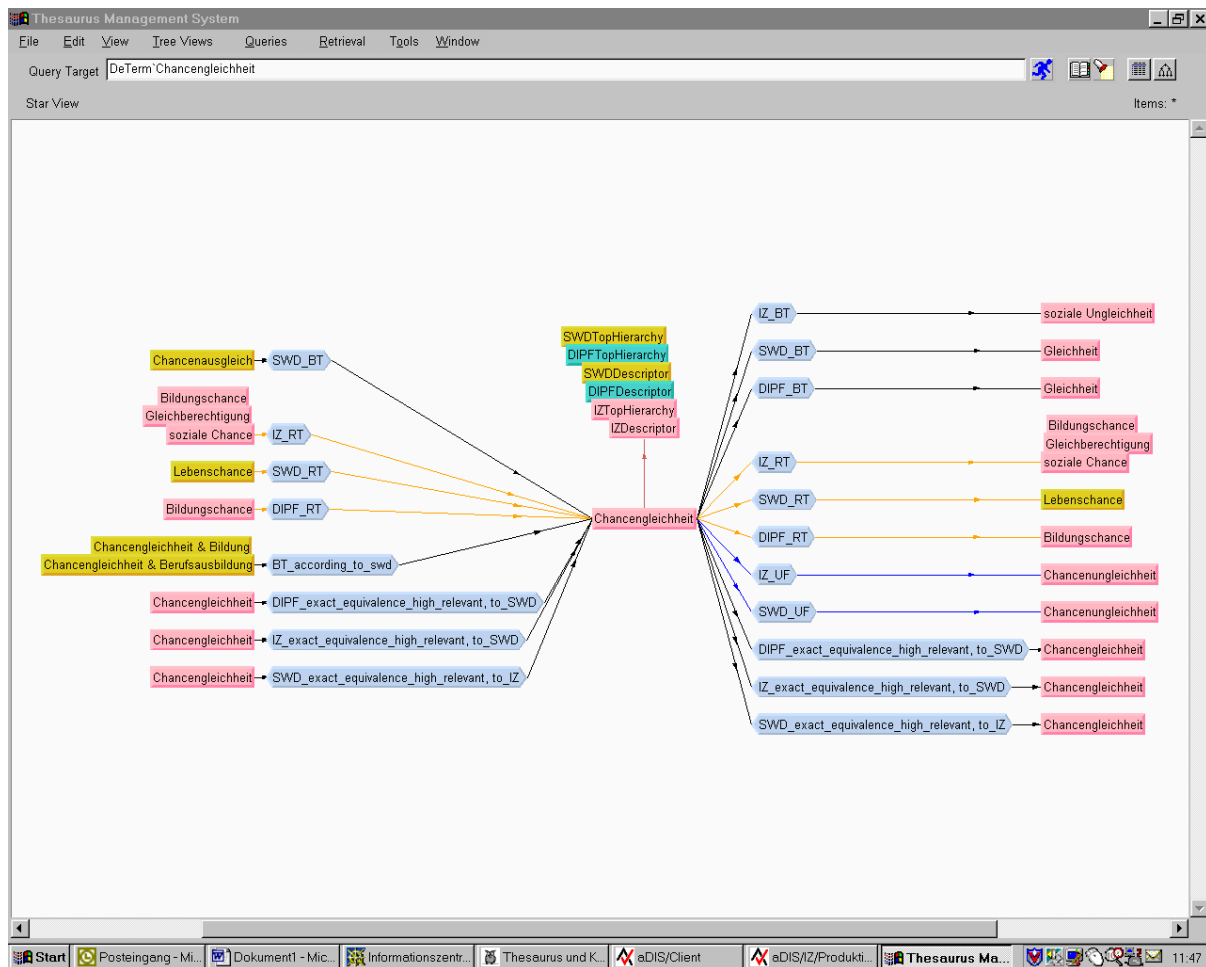
rot = equivalence

gelb = related term

blau = narrower -> broader term

b) aus der Anwendung in CARMEN AP 12

das Schlagwort „Chancengleichheit“ in der „Star View“ von SIS-TMS



- „Chancengleichheit“ ist in allen drei Thesauri ein Deskriptor (zugelassenes Schlagwort);
- als Konkordanzrelation wurde in allen drei Fällen „exact_equivalence_high_relevant“ vergeben;
- „Chancengleichheit“ hat innerhalb des jeweiligen Thesaurus auch jeweils einen Verweis (schwarzer Pfeil) auf einen broader term (BT) - „soziale Ungleichheit“ beim TheSoz; „Gleichheit“ bei ThesBild und SWD;
- „Chancengleichheit“ zeigt innerhalb des jeweiligen Thesaurus auch mit einem gelben Pfeil auf related terms (RT) - „Bildungschance/ Gleichberechtigung/ soziale Chance“ beim TheSoz, „Lebenschance“ bei der SWD und „Bildungschance“ beim ThesBild

SIS-TMS wurde für die Präsentation und Bearbeitung der Crosskonkordanzen von Thesauri in Kooperation mit AP 11 angepasst und in AP 12 zum Einsatz gebracht. In der Praxis zeigte sich, dass die grafische Oberfläche gute Dienste leistet, wenn man sich über bereits vorhandene Relationen innerhalb oder zwischen Thesauri informieren will. Wenig benutzerfreundlich ist das Eingabetool zur Bearbeitung von Begriffen und Relationen (mehrstufige, nicht selbsterklärende Dialogfolge beim Einstieg und der Bearbeitung, keine Durchlässigkeit zwischen Thesauri und Konkordanzbildungen für Drag&Drop-Aktionen). Da Verbesserungen an

SIS-TMS während der Projektlaufzeit zeitlich nicht möglich waren, wurden die Konkordanzen zunächst als Excel-Tabellen erstellt und dann mittels in Java programmierten Parsern in SIS-TMS eingelesen und auf dem Server auch für externe Zugriffe bereitgestellt. Inzwischen wurde seitens des IZ an ICS-FORTH ein Auftrag zur Verbesserung von SIS-TMS erteilt.

1.3 Crosskonkordanzen von Klassifikationen (UAP 4, 7) RVK, BK⁴, IZK, MSC, PACS

Die Ausgangsbasis für die Crosskonkordanzen bildeten die bibliothekarischen Allgemein-klassifikationen Regensburger Verbundklassifikation (RVK) und Basisklassifikation (BK) sowie die Fachklassifikationen Mathematics Subject Classification (MSC), Physics and Astronomy Classification Scheme (PACS) und die Klassifikation Sozialwissenschaften des IZ (IZK). Die Untersuchung hat zwei fachliche Schwerpunkte, einmal die Fächer Mathematik und Physik, und zum anderen den Bereich Sozialwissenschaften. Die Crosskonkordanzen im Bereich Sozialwissenschaften wurden vom Informationszentrum Sozialwissenschaften erstellt, Mathematik und Physik von der Universität Regensburg.

Mathematik und Physik, Mengengerüst

Klassifikation	Anzahl der Systemstellen		
	gesamt	Mathematik	Physik
RVK	~ 800 000	62	1 146
BK	2 128	44	49
PACS (1999)			3 370
MSC		5 514	

Sozialwissenschaften, Mengengerüst

Klassifikation	Anzahl der Systemstellen	
	gesamt	Sozialwissenschaften
RVK	~ 800 000	~ 2 520
BK	2 128	~ 320
IZ		159

⁴ die Basisklassifikation (BK) wird z.B. vom Gemeinsamen Bibliotheksverbund mit der Software PICA genutzt (daher die verkürzte Bezeichnung GBV/PICA im Projektantrag)

1.3.1 Die Ergebnisse im Überblick

Das folgende Diagramm zeigt die im Projekt erstellten Crosskonkordanzen.

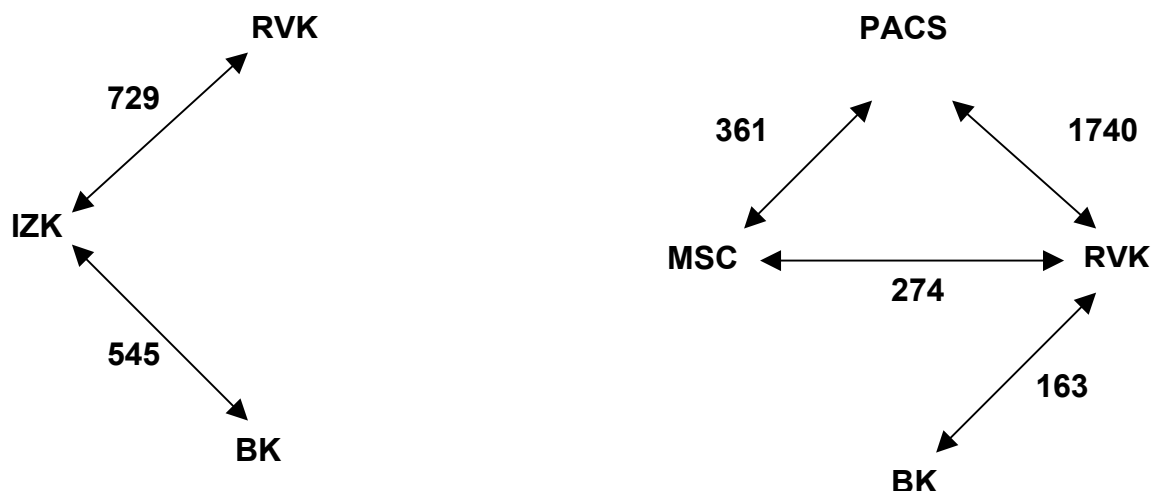


Abbildung: Crosskonkordanzen von Klassifikationen

Die Zahlen stehen für die Anzahl der erstellten Verknüpfungen, links der Bereich Sozialwissenschaften, rechts die Fächer Physik und Mathematik.

1.3.2 Crosskonkordanzen von Klassifikationen im Bereich Sozialwissenschaften

1.3.2.1 Klärung inhaltlich/ methodischer Fragen

Eine grundsätzliche Schwierigkeit bestand in der verschiedenen Historie bzw. "Philosophie" der Regensburger Verbundklassifikation (RVK) und der Klassifikation Sozialwissenschaften (IZK) bzw. der Basisklassifikation (BK).

Die RVK ist aus einer universal ausgerichteten Bibliothekssystematik entstanden. Eine solche muss einerseits alle Fachgebiete abdecken, andererseits aber auch eine standortbezogene Feingliederung zur Aufstellung von Büchern und Zeitschriften bieten. Die IZK hingegen wurde als Fachklassifikation zur Recherche in Online-Datenbanken entwickelt, um zusätzlich zur Suche mit Textwörtern oder kontrollierten Schlagwörtern ein weiteres Instrument zur Verfügung zu haben, mit dem sich Gruppen von Literatur- oder Projektnachweisen bei einer Suchanfrage ein- oder ausschließen lassen. Die IZK umfasst alle sozialwissenschaftlichen Fachgebiete, ist hierarchisch nicht tief gestaffelt, sondern an den in den wissenschaftlichen Fächern häufig anzutreffenden Differenzierungen orientiert (z.B. spezielle Soziologien wie im Lehrangebot der Universitäten). Die an Stelle der DDC bearbeitete Basisklassifikation (BK) besitzt ebenfalls eine standortunabhängige Fachsystematik und ist insofern der IZK ähnlicher als der RVK. Andererseits deckt sie mehr oder weniger alle Fachgebiete ab, hat also einen universellen Charakter wie die RVK. Dennoch ist die Systematik an vielen Stellen anders gegliedert als die RVK oder die IZK, was in Einzelfällen die Zuordnung schwierig oder unmöglich macht. So werden z.B. bei der IZK „Pädagogik“ und „Bildungswesen“ unter die Hauptklasse „Erziehungswissenschaft“ subsumiert, während die entsprechenden Bereiche in der BK getrennte Hauptklassen bilden. In ihrer relativ flachen Struktur und ihrer Funktionalität ähnelt die BK jedoch der IZK, d.h. sie ist zur Feinrecherche in großen Katalogen bzw. Datenbanken

nur bedingt geeignet. Insgesamt stellte sich der Aufbau der Crosskonkordanz BK-IZK – wie auch erwartet – einfacher dar als die Anlage der Crosskonkordanz RVK-IZK.

1.3.2.2 Crosskonkordanzen RVK - IZK (UAP 7)

Für die erste Phase der intellektuellen Erarbeitung der Crosskonkordanzen wurden entsprechend dem Projektantrag die Fachgebiete Soziologie, Methoden der Sozialforschung, Politikwissenschaft und Bildungsforschung ausgewählt und mit Hilfe des von der Universitätsbibliothek Regensburg entwickelten Programms CarmenX online bearbeitet. Die zulässigen Typen von Relationen waren

- Äquivalenz („=“)
- Oberbegriff („>“)
- Unterbegriff („<“)
- verwandter Begriff („><“)

Die Intensität der Beziehung konnte mit "hoch", "mittel" und "gering" angegeben werden. Diese zu Beginn festgelegten Kriterien erwiesen sich im großen und ganzen als geeignet und brauchten im Laufe der Arbeit nicht geändert zu werden. Wegen der unterschiedlichen Struktur der beiden Klassifikationen wurde die Relation "äquivalent" jedoch äußerst selten vergeben. Die drei Intensitätsmaße erwiesen sich als voll ausreichend, um die Stärke der jeweiligen Beziehung zu kennzeichnen.

Softwaretechnische Arbeitsgrundlage zur Festlegung der Konkordanzbeziehungen war CarmenX der UB Regensburg, das als Ausgangspunkt die Klassifikationen als zweispaltige Dateien bereitstellte – zum einen für die Verknüpfungen vom RVK zur IZK und zum anderen für die gegenläufigen Verknüpfungen von der IZK zur RVK. Ausgangspunkt für die intellektuelle Bearbeitung war – wegen der besseren Kenntnis der Systematik und der Klassifizierungspraxis – die IZK. Innerhalb der IZK wurden ca. 100 Klassen ausgewählt, die zunächst einseitig mit ca. 500 Klassen der RVK verknüpft wurden. In einem weiteren Arbeitsschritt wurden die einschlägigen Klassen der RVK darauf untersucht, ob sie von Seiten der IZK bereits einen Konkordanzeintrag erhalten hatten. Traf dies zu, wurden sie aus der Sicht der RVK noch einmal beleuchtet und ggf. geändert. Außerdem wurden weitere Konkordanzen aufgebaut und die gegenläufigen Verknüpfungen IZK-RVK erstellt. Hierbei zeigte sich, dass die Beziehungen fast immer symmetrisch waren. Bei Unklarheiten über die Inhalte bzw. die Definition der einzelnen Klassen wurde der Bayerische Verbundkatalog (BVB) bzw. die Literaturdatenbank des Informationszentrums Sozialwissenschaften (SOLIS) zu Rate gezogen.

Insgesamt bleibt festzuhalten, dass ca. 1000 Verknüpfungen manuell erstellt und mehrfach – theoretisch und empirisch – überprüft wurden. Hinzu kam eine gewisse Anzahl von Klassen, die zwar einschlägig waren, zu denen aber kein passendes Gegenstück gefunden werden konnte.

Beispiel: Wechselseitige Verknüpfung zwischen einer RVK- und einer IZK-Klasse
(MS 1200-MS 1277 \leftrightarrow 10203)

The screenshot shows the X-Walk web interface in a Netscape browser window. The address bar shows the URL: <http://rzb4.bibliothek.uni-regensburg.de/xwalk/xwalk.html?k1=rvko&k2=iz>. The interface is divided into two main panes. The left pane shows a hierarchical tree structure of classification classes. The right pane shows a detailed view of a specific relationship between two classes.

Left Pane (Classification Tree):

- ▲ MN - MW Soziologie
 - iz: [10200](#)
 - ▲ MS Spezielle Soziologien
 - ▼ MS 1000 - MS 1060 Kleingruppen
 - iz: [10706](#)
 - ▼ MS 1090 - MS 1190 Gesamtgesellschaften
 - iz: [10203](#)
 - ▼ MS 1200 - MS 1277 Analysen einzelner Länder
 - iz: [10203](#)
 - MS 1280 Theorien und Probleme der Massengesellschaft
 - iz: [10203](#)

Right Pane (Detailed View):

Makrosoziologie, spezielle Theorien und "Schulen", Entwicklung und Geschichte der Soziologie
 rvko: [MP 1000](#) [MP 3000](#) [MP 5000](#) [MQ](#) [MR 4000](#) [MR 5100](#) - [MR 5800](#) [MR 7100](#) - [MR 7400](#) [MS 1300](#) - [MS 1480](#) [MS 1500](#) - [MS 1560](#) [ZB 52100](#) [ZB 52300](#) [ZB 52550](#) [MP](#)

10202 Lehre und Studium, Professionalisierung und Ethik, Organisationen und Verbände der Soziologie
 rvko: [AL 41500](#) [MP 6000](#) [MP 8000](#)

► 10203 Soziologie von Gesamtgesellschaften
 rvko: [BL 5200](#) - [BL 5460](#) [CC 7750](#) [MS 1090](#) - [MS 1190](#) [MS 1200](#) - [MS 1277](#) [MS 1280](#) [ZB 54600](#) [ZB 54000](#)

bestehende Beziehung (wechselseitig)

rvko	Beziehung	iz	Relevanz	Bemerkung	Ändern	Löschen
MS 1200 - MS 1277	<	10203	hoch			

Dokument: Übermittelt

In der zweiten Phase des intellektuellen Abgleichs der IZK mit der RVK wurden die weiteren Fachklassen der IZK mit der RVK verknüpft, alle Relationen noch einmal überarbeitet und exemplarisch anhand der Datenbankeinträge im Bibliotheksverbund Bayern (BVB) und der Literaturdatenbank des Informationszentrums Sozialwissenschaften (SOLIS) überprüft. Bei diesem Abgleich zeigten sich die gleichen Probleme wie bei Phase 1: Wegen der verschiedenen Struktur der beiden Klassifikationen war es häufig schwierig, eine passgenaue Verknüpfung zu finden. Es mussten daher nicht selten - ausgehend von einer Klasse auf der einen Seite - Verknüpfungen zu mehreren Klassen auf der anderen Seite hergestellt werden. Hieraus erklärt sich auch die relativ große Zahl von ca. 1.450 Verknüpfungen insgesamt. Die Äquivalenz-Relation (=) wurde auch in der zweiten Phase sehr selten vergeben. Dagegen ist die Relation „enger : weiter“ (<) viel häufiger vertreten als „weiter : enger“ (>), was darauf hinweist, dass die RVK viel tiefer durchgegliedert ist als die IZK. Damit wurden die bereits in Phase 1 gemachten Feststellungen bestätigt. Die Ähnlichkeitsrelation (><) wurde ebenfalls relativ selten vergeben. Die beigefügte Übersicht informiert über die Zahl und den Typus der hergestellten Verknüpfungen RVK - IZK im einzelnen (da die Verknüpfungen fast immer spiegelbildlich sind, wird auf die Auflistung der gegenläufigen Relationen verzichtet).

Übersicht: Häufigkeit der Relationen zwischen RVK und IZK

Typ der Relation	Anzahl	Relevanz	Zwischensummen
=	30	gleich	30
<	294	hoch	528
<	120	mittel	
<	114	niedrig	
>	15	hoch	37
>	19	mittel	
>	3	niedrig	
><	49	hoch	134
><	57	mittel	
><	28	niedrig	
gesamt	729		729

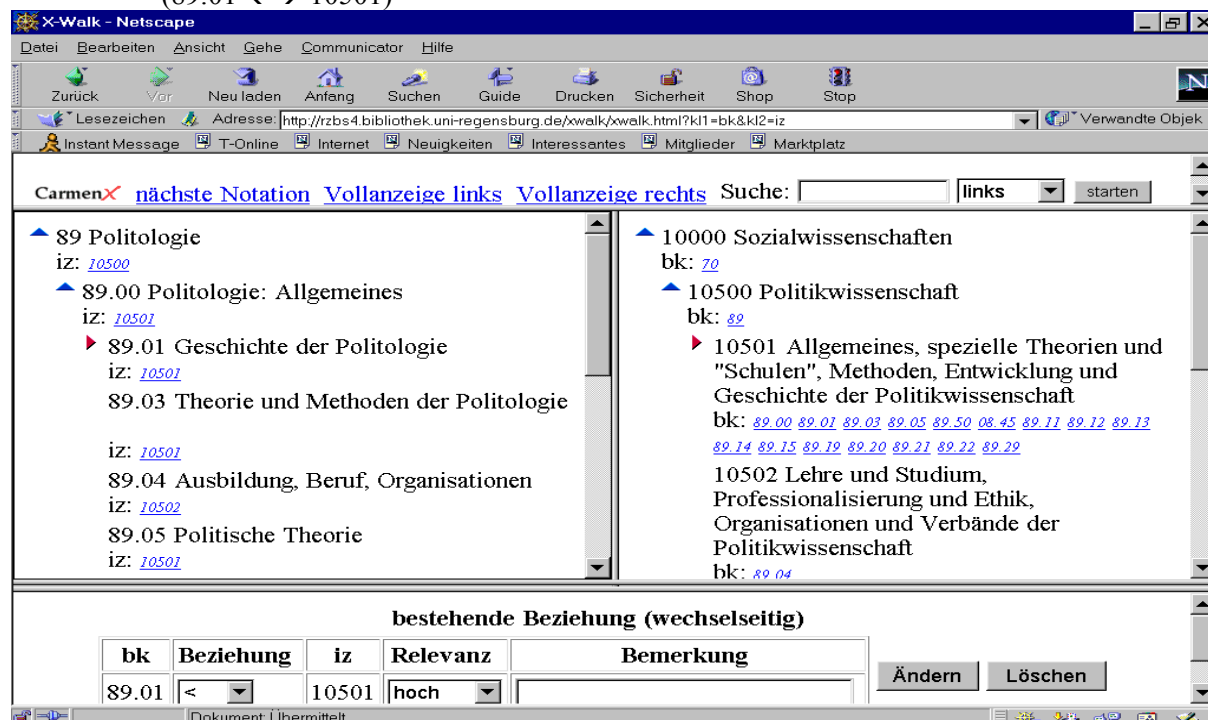
1.3.2.3 Crosskonkordanzen BK [GBV/PICA] - IZK (UAP 9)

Da die DDC wegen Copyrightproblemen nicht zum Aufbau von Crosskonkordanzen zur Verfügung stand, wurden die Crosskonkordanzen BK – IZK nicht nur, wie im Projektantrag vorgesehen, für vier Fachgebiete erstellt, sondern für alle Klassen der IZK. Die Bearbeitung erfolgte ebenfalls online mit Hilfe des Programms CarmenX, dessen Funktionsumfang sich auch hier als voll ausreichend erwies.

Die Konkordanzen wurden sowohl in der Richtung BK → IZK wie auch in umgekehrter Richtung IZK → BK erstellt.

In Zweifelsfällen wurde der Inhalt der einzelnen Klassen anhand des Verbundkataloges des GBV bzw. des Darmstädter Verbundkataloges (DVK) sowie der Literaturlatenbank SOLIS überprüft.

Beispiel: Wechselseitige Verknüpfung zwischen einer BK- und einer IZ-Klasse (89.01 \leftrightarrow 10501)



Ergebnis: Mit knapp 1.100 Relationen insgesamt ergaben sich weniger Verknüpfungen als zwischen RVK und IZK (1.450). Ursache hierfür war - wie schon oben dargelegt - eine flachere und im ganzen auch ähnlichere Struktur von BK und IZK, die weniger Mehrfachverknüpfungen erforderte. Die Verteilung der verschiedenen Verknüpfungstypen ist dennoch bei beiden Paarungen ähnlich. Die nachfolgende Übersicht informiert über die Zahl und den Typus der hergestellten Verknüpfungen BK - IZK im einzelnen (da die Verknüpfungen fast immer spiegelbildlich sind, wird auf die Auflistung der gegenläufigen Relationen verzichtet).

Übersicht: Häufigkeit der Relationen zwischen BK und IZK

Typ der Relation	Anzahl	Relevanz	Zwischensummen
=	20	gleich	20
<	292	hoch	394
<	59	mittel	
<	43	niedrig	
>	12	hoch	33
>	18	mittel	
>	3	niedrig	
><	42	hoch	98
><	38	mittel	
><	18	niedrig	
Gesamt	545		545

Die Zahl der Äquivalenzen ist niedrig, der Typus „enger : weiter (<)“, verbunden mit hoher Relevanz, bei weitem der häufigste. Dies liegt daran, dass die Struktur der BK immer noch tiefer durchgegliedert ist als die der IZK.

1.3.2.4 Exemplarische Test der Crosskonkordanzen anhand von Datenbanken

In einer nachfolgenden Testphase wurden die hergestellten Konkordanzen zwischen RVK und IZK anhand der Datenbanken des Bibliotheksverbundes Bayern (BVB) und der Literaturdatenbank SOLIS des IZ Sozialwissenschaften überprüft (Ergebnisse siehe nachfolgende Übersicht). Wegen der großen Zahl der Relationen und der noch größeren Zahl der Dokumente in den beiden Datenbanken BVB und SOLIS konnten die Tests nur exemplarisch vorgenommen werden. Dennoch scheint die Folgerung zulässig, dass die meisten Relationen relativ treffsicher sind und die beiden Klassifikationen auch in der Praxis verbunden werden können.

Übersicht: Exemplarischer Test der unter CARMEN X angelegten wechselseitigen Verknüpfungen sozialwissenschaftlicher Klassifikationen zwischen der Regensburger Verbundklassifikation (RVK) und der IZ- Klassifikation (IZK) anhand des Bibliotheksverbundes Bayern (BVB) und der Literaturdatenbank des Informationszentrum Sozialwissenschaften (SOLIS)

Regensburger Verbundklassifikation		Trefferzahl BVB		IZ Klassifikation		Trefferzahl SOLIS	
Notation	Text		Relation (Relevanz)	Notation	Text		Bewertung ⁵
DS	Sozialpädagogik, Sozialarbeit	5358	= (hoch)	20600	Sozialarbeit und Sozialpädagogik	5093	++
MS 6530	Soziologie der Freizeit	756	< (hoch)	20400	Freizeitforschung, Freizeitsoziologie	3408	++
MR 2000 - MR 2950	Methoden der Sozialforschung	16348	> (hoch)	10104	Forschungsarten der Sozialforschung	2583	+ (IZ- Klasse enger definiert, daher nicht "=""
ME	Politische Systeme: Allgemeines und Vergleich	3794	>< (hoch)	10503	Staat, politisches System	3319	+ (Schnittmenge beträgt ca. 75 %)
DF	Systematische Pädagogik	7422	< (mittel)	10601	Allgemeines, spezielle Th. u. „Schulen“, Meth., Entwickl. und Gesch. der Erziehungswissenschaft	558	++
AN	Buch- und Bibliothekswesen, Informationswissenschaft	18810	> (mittel)	1080500	Informationswissenschaft	83	+ (Anteil der Arbeiten zur Informationswissenschaft beim BVB gering)
MB 3250	Politik und Soziologie	208	>< (mittel)	10504	politische Willensbildung, politische Soziologie		+ (Schnittmenge relativ groß, Beziehung könnte aber auch als "<" definiert werden)
PG	Rechtspflege, Gerichtsverfassung, Zivilverfahrensrecht und Nebengebiete	6396	< (niedrig)	40102	Justiz	571	++
XF 1218	Medizinische Ausbildung: Medizinische Psychologie, medizinische Soziologie	58	> (niedrig)	10708	psychische Störungen, Behandlung und Prävention	2106	-- (BVB enthält weitgehend Lehrbücher zur ärztlichen Vorprüfung, SOLIS Arbeiten zu den sozialen Folgen psychischer Störungen, daher "><" besser geeignet)
AL 41500	Hochschul- und Universitätswesen: Wirtschafts-, Sozial- und Politische Wissenschaften	85	>< (niedrig)	10202	Lehre und Studium, Professionalisierung und Ethik, Organisationen und Verbände der Soziologie	222	+ (beide Klassen enthalten Studienführer, SOLIS zusätzlich Arbeiten zur Entwicklung und Situation der einzelnen Lehrfächer: evtl. in "<" ändern)

⁵ Stichprobe n=50 auf beiden Seiten (jeweils neueste Einträge), ++ = voll zutreffend, + = weitgehend zutreffend, -= teilweise zutreffend, -- = gar nicht zutreffend

Der Test der zwischen der BK und der IZK angelegten Crosskonkordanzen erfolgte anhand des Darmstädter Verbundkataloges und der Datenbank SOLIS. Der Test war aufgrund von nicht angekündigten Umstellungen am DVK, die zeitlich in die Testphase fielen, und aufgrund der geringen Besetzung mancher Klassen sehr schwierig. Als Ergebnis (vgl. dazu Übersicht 4) kann jedoch - ähnlich wie beim Abgleich der Einträge zwischen BVB und SOLIS - festgehalten werden, dass die Verbindung BK-IZ tragfähig und ein Übergang von einem System in das andere durchaus möglich ist.

Übersicht: Exemplarischer Test der unter CARMEN X angelegten wechselseitigen Verknüpfungen sozialwissenschaftlicher Klassifikationen zwischen der Basisklassifikation (BK) und der IZ- Klassifikation (IZK) anhand des Darmstädter Verbundkataloges (DVK) und der Literaturdatenbank des Informationszentrum Sozialwissenschaften (SOLIS)

Basisklassifikation		Trefferzahl DVK		IZ Klassifikation		Trefferzahl SOLIS	
Notation	Text		Relation (Relevanz)	Notation	Text		Bewertung ⁶
71.35	Kindersozio- logie, Jugendsoziologie	161	= (hoch)	10210	Jugendsoziologie	8757	++
71.00	Soziologie: Allgemeines	65	< (hoch)	10201	Allg. Soziologie, Makroso- ziologie, spez. Theorien u. "Schulen", Entwicklung u. Gesch. d. Soziologie	12106	++ (IZ-Klasse erheblich anders definiert, dennoch ist auch das Spektrum des DVK enthalten)
70.03	Methoden, Techniken und Organisation der sozialwissenschaftlichen Forschung	116	> (hoch)	10104	Forschungsarten der Sozialforschung	2601	++
71.11	Gesellschaft	98	>< (hoch)	10203	Soziologie von Gesamtgesellschaften	3315	++
15.06	Politische Geschichte	98	< (mittel)	30301	allgem. Geschichte	3222	+ (Inhalt der jeweiligen Klassen sehr ähnlich, daher hätte auch "><" vergeben werden können)
85.05	Betriebssoziologie, Betriebspsychologie	136	>...(mittel)	10205	Wirtschaftssoziologie	9102	- (relativ wenige Überschneidungen zwischen den Klassen, daher besser >< (niedrig)
05.20	Kommunikation und Gesellschaft	115	>< (mittel)	1080400	Massenkommunikation	1285	++
02.00	Wissenschaft und Kultur allgemein	47	<... (niedrig)	10102	Wissenschaftstheorie, Wissenschaftsphilosophie, Wissenschaftslogik	4902	+ +
71.80	Sozialpolitik: Allgemeines	103	>...(niedrig)	11002	Lehre und Studium, Professionalis. und Ethik, Organisationen u. Verbände der Sozialpolitik	38	++
83.66	Agrarwirtschaft	46	>< (niedrig)	10212	Agrarsoziologie	1884	++

⁶ Stichprobe n=50 (jeweils neueste Einträge), außer in drei Fällen wegen zu geringer Gesamtstrefferzahl, ++ = voll zutreffend, + = weitgehend zutreffend, - = teilweise zutreffend, -- = nicht zutreffend

1.3.3 Crosskonkordanzen von Klassifikationen für die Fächer Physik und Mathematik

1.3.3.1 Klärung inhaltlicher/methodischer Fragen

Mathematics Subject Classification (MSC) wurde von www.ams.org in der aktuellen Version MSC2000, Physics and Astronomy Classification Scheme (PACS) wurde in der Version 1999 und 2001 von www.aip.org aus dem Internet bezogen. Für die Crosskonkordanzen wurde die PACS 1999 verwendet. Die Basisklassifikation wurde in der 3. erweiterten Auflage von 2000 vom Gemeinsamen Bibliotheks Verbund (GBV) (www.gbv.de) übernommen.

Crosskonkordanzbeziehungen nach Typ und Relevanz

	PACS → RVK	MSC → RVK	RVK → BK	MSC → PACS
Zahl der Verknüpfungen	1.935	274	163	361
Äquivalenz (=)	26%	32%	42%	39%
Oberbegriff (<)	14%	17%	14%	14%
hohe Relevanz	11%	13%	12%	12%
mittlere Relevanz	3%	4%	2%	2%
geringe Relevanz	0%	0%	--	--
Unterbegriff (>)	33%	36%	33%	31%
hohe Relevanz	16%	25%	5%	24%
mittlere Relevanz	15%	10%	17%	6%
geringe Relevanz	3%	1%	12%	2%
Verwandter Begriff (><)	27%	15%	11%	16%
hohe Relevanz	13%	5%	9%	12%
mittlere Relevanz	12%	9%	2%	4%
geringe Relevanz	2%	0%	1%	--

1.3.3.2 Crosskonkordanz PACS – RVK

PACS und RVK lassen sich recht gut aufeinander abbilden. Der Umfang der Systemstellen ist etwa gleich, wenn man berücksichtigt, dass die PACS über die Physik hinaus auch viele Bereiche angrenzender Wissenschaften enthält. Etwa 22% der Verknüpfungen aus der PACS heraus gehen zu anderen Fachsystematiken der RVK. Die Strukturen beider Klassifikationen sind sich recht ähnlich. Die Ähnlichkeit beruht darauf, dass man sich bei der Entwicklung der Fachsystematik Physik an der damaligen Version der PACS orientiert hatte.

1.3.3.3 Crosskonkordanz RVK – MSC

Eine Crosskonkordanz zwischen RVK und MSC ist durch die stark unterschiedliche Gliederungstiefe geprägt. Die Fachsystematik Mathematik der RVK zeichnet sich durch einen hohen Anteil formaler Systemstellen aus. Es gibt nur 62 sachliche Systemstellen, denen 5514 Stellen in der MSC gegenüberstehen.

1.3.3.4 Crosskonkordanz RVK – BK

Zwischen RVK und BK wurde eine Crosskonkordanz für Physik und Mathematik erstellt. Der begrenzende Faktor ist auch hier die geringe Gliederungstiefe der Basisklassifikation. Mit nur 163 Konkordanzbeziehungen für beide Fächer sind große Treffermengen bei der Suche unvermeidlich. Die Mathematik-Systematik der RVK hat neben inhaltlichen viele formale Systemstellen wie etwa Schriftenreihen. Dies entspricht ihrem Ursprung als Aufstellungssystematik. Die Basisklassifikation

hat keine vergleichbare Gliederung nach formalen Kriterien. Eine Konkordanz zwischen RVK und BK im Bereich Mathematik ist sinnvoll, da die beiden Teilbereiche ähnlich tief gegliedert sind. Für den Teilbereich Physik ist die RVK sehr viel stärker untergliedert als die Basisklassifikation. Trotzdem ist dort eine relativ gute Übereinstimmung mit den (Physik-)Teilgebieten feststellbar.

1.3.3.5 Crosskonkordanz PACS – MSC

Die Crosskonkordanz zwischen PACS und MSC unterscheidet sich von allen anderen in diesem Projekt aufgestellten Crosskonkordanzen. Hier ist der Versuch gemacht, zwei verschiedene Fächer miteinander in Beziehung zu setzen. PACS und MSC gleichen sich darin, dass sich in jeder Fachsystematik das jeweils andere Fach in einer Art Grobssystematik wiederfindet. Die PACS hat einen mathematischen Teil und die MSC einen physikalischen Teil. Die hier erstellte Crosskonkordanz bildet genau diesen Zusammenhang ab. Nicht abgebildet wird der Zusammenhang zwischen einem physikalischen Gegenstand und einer dazu passenden mathematischen Methode. Dafür ist ein intellektuelles Verfahren ungeeignet.

1.3.3.6 Bewertung der Ergebnisse

a) Vergleich zwischen intellektuellem (AP 12) und empirisch/statistischem Verfahren (AP 9)

In Arbeitspaket AP 9 wurde mit einem empirisch/statistischen Verfahren eine Crosskonkordanz zwischen MSC und PACS erstellt. Der Vergleich zeigt, dass die Ergebnisse sich grundsätzlich voneinander unterscheiden. Auffällig ist, dass beim Verfahren in AP 9 viel mehr Verknüpfungen angelegt werden, als beim intellektuellen Verfahren. Beim intellektuellen Verfahren wird überwiegend eine Verbindung zwischen einer Systemstelle aus PACS und MSC hergestellt. Beim empirisch/statistischen Verfahren erhält man eine Vielzahl von Verbindungen. Z.B. ist 03.65-w Quantenmechanik aus PACS mit 205 Systemstellen der MSC verbunden. Die Unterschiede erklären sich dadurch, dass das empirisch/statistische Verfahren einen anderen Zusammenhang zwischen Physik und Mathematik herstellt als das intellektuelle Verfahren. Beim statistisch/empirische Verfahren erhält man zu einem physikalischen Gegenstand die verwendete mathematische Methode. Beide Methoden stehen also hier nicht in Konkurrenz zueinander, sondern ergänzen sich.

Übersicht: Exemplarische Tests der Crosskonkordanzen zwischen Regensburger Verbundklassifikation (RVK) und Basisklassifikation (BK) durch Testrecherchen im Katalog des Bibliotheksverbundes Bayern (BVB) und des Gemeinsamen Bibliotheksverbundes (GBV).

Regensburger Verbundklassifikation		Trefferzahl BVB		Basisklassifikation		Trefferzahl GBV	
Notation	Text		Relation (Relevanz)	Notation	Text		Bewertung ⁷
UK 1000	Quantentheorie, allgemein Nichtrelativistische Quantentheorie	527	= (hoch)	33.23	Quantenphysik	1400	++
UN	Kernphysik	1976	= (hoch)	33.40	Kernphysik	1878	++
UM	Atom- und Molekülphysik	1339	= (hoch)	33.30	Atomphysik, Molekülphysik	902	++
UO 4000 - UO 4100	Feldtheorie und Quantenfeldtheorie	443	>< (hoch)	33.24	Quantenfeldtheorie	610	+ (In der RVK teilweise unter UK anzusiedeln)
UX	Experimentelle Methoden der Physik und Physikalische Technologie	2712	>< (hoch)	33.05	Experimentalphysik	1093	- (In RVK/BVB mehr die rein technische Seite – Geräte, Elektronik, in BK/GBV mehr physikalische Sichtweise)
SN 300	Unterhaltungsmathematik, mathematische Spiele	522	< (hoch)	31.08	Unterhaltungsmathematik	373	+ (In der RVK stärker unterschieden)
SK 380	Klassische Geometrie	1736	> (mittel)	31.51	Algebraische Geometrie	1020	+
SG 700	Philosophie und Mathematik	414	< (hoch)	31.02	Philosophie und Wissenschaftstheorie der Mathematik	1213	+

⁷ Stichprobe n=50 (jeweils neueste Einträge), ++ = voll zutreffend, + = weitgehend zutreffend, - = teilweise zutreffend, -- = nicht zutreffend

Übersicht: Exemplarische Tests der Crosskonkordanzen zwischen Regensburger Verbundklassifikation (RVK) und Physics and Astronomy Classification Scheme (PACS) durch Testrecherchen im Katalog des Bibliotheksverbundes Bayern (BVB) und der Fachdatenbank INSPEC.

Regensburger Verbundklassifikation		Treffer BVB		PACS		Treffer INSPEC	
Notation	Text		Relation (Relevanz)	Notation	Text		Bewertung ⁸
UH 8200	Spezielle Relativitätstheorie	127	= (hoch)	03.30+p	Special relativity	4586	+ Relativität manchmal nur ein Seitenaspekt. Möglichkeit der Vergabe mehrerer Notationen in INSPEC
UH 5630	Gaslaser, Ionenlaser	45	< (hoch)	42.55.Lt	Gas lasers including excimer and metal-vapor lasers	2973	+ In der neuen Ausgabe der PACS sind Gas- und Excimer-Laser getrennt. Hierfür müsste dann bei der neuen Notation Gleichheit (=) auftreten.
UN 1900	Radioaktivität allgemein	185	= (hoch)	23.	Radioactive decay and in- beam spectroscopy	31	++ aber Ergebnisse in INSPEC viel spezieller. Liegt an den beschriebenen Dokumenten (Bücher ↔ Artikel). PACS hat auch einen ganzen Hierarchie-Baum, RVK nur eine einzige Notation.
UN 1200	(Kernstruktur /)Allgemeines	38	> (mittel) (abgeleitet)	21.10.Dr	Binding energies and masses	5185	+ (allgemein ↔ z.T. sehr speziell, ist aber klar)
UK 1250	Theorie des Messprozesses	23	< (hoch)	03.65.Bz	Foundations, theory of measurement, miscellaneous theories (including Aharonov Bohm effect, Bell inequalities, Berry's phase)	12631	+ (in diesem Zusammenhang viel verschränkte Zustände, Bell'sche Ungleichungen)
UK 3500	Drehimpulse in der Quantenmechanik	32	< (hoch)	03.65.Fd	(Quantum mechanics/) Algebraic methods	8807	+ (Das Gewicht ist auf jeden Fall zu erniedrigen → mittel)
US 2000	Astrophysik Kosmologie und Kosmogonie allgemein	416	> (hoch)	95.30.-k	Fundamental aspects of astrophysics	55998	++
US 3100 - US 3600	Galaxien	190	= (hoch)	98.	Stellar systems; interstellar medium; galactic and extragalactic objects and systems; the Universe	137699	++
UG 3800	Theorie der Phasenübergänge	184	> hoch)	64.60.-i	General studies of phase transitions	14446	- (Bei den Treffern in INSPEC stand nicht immer der Phasenübergang im Vordergrund, sondern war oft nur Nebenaspekt.)

⁸ Stichprobe n=50 (jeweils neueste Einträge), ++ = voll zutreffend, + = weitgehend zutreffend, - = teilweise zutreffend, -- = nicht zutreffend

1.4 Crosskonkordanzen von Thesauri (UAP 8, 13, 15)

Schlagwortnormdatei (SWD), Thesaurus Sozialwissenschaften (TheSoz), Schlagwortmaterial des FIS Bildung (ThesBild)

Die Crosskonkordanzen zwischen der SWD an Der Deutschen Bibliothek (DDB) und dem TheSoz des Informationszentrums Sozialwissenschaften (IZ) sowie dem Schlagwortmaterial des FIS Bildung (ThesBild) des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) wurden in enger Kooperation zwischen dem IZ, dem DIPF und DDB erstellt.

1.4.1 Klärung inhaltlich/methodischer Fragen

a) Struktur der Thesauri

Die SWD ist ein Universalthesaurus mit ca. 142.100 Sachschlagwörtern (Stand Jan. 2002), in dem laufend Ergänzungen und Änderungen wirksam werden. Über Notationen sind die Schlagwörter systematisch geordnet. Der TheSoz ist mit 11.100 Einträgen (Stand Jan. 2002) ein umfassender Fachthesaurus für die Sozialwissenschaften. Ein Update erfolgt jährlich, eine systematische Ordnung ist ebenfalls über Notationen vorhanden. ThesBild besteht aus insgesamt 60.000 Schlagwörtern von denen ca. ein Drittel Sachschlagwörter darstellen. Die Termini sind über Notationen der Fachsystematik Bildung rubriziert. In allen drei Thesauri sind die klassischen Thesaurusrelationen Äquivalenz, Hierarchie und Verwandtschaft ausgewiesen.

Die Strukturprinzipien zeigen also eine relativ hohe Übereinstimmung, die die Verknüpfung der Thesauri erleichtert.

Unterschiede bestehen in folgenden Punkten:

- unterschiedlich ausgeprägtes Maß an Präkoordination der Begriffe
- in der kleineren Menge an Eigennamen im TheSoz
- in der Kennzeichnung von Homonymen in der SWD und im ThesBild, TheSoz arbeitet hier mit Scope Notes und use-Verweisen;
- in der Unterscheidung von Schlagwortkategorien (Sachschlagwörter, Personenschlagwörter, Körperschaftsschlagwörter, Forms Schlagwörter, Geographika) in der SWD. Für die Konkordanzbildung wurden die Sachschlagwörter genutzt.

Die Indexierungsregeln und -konventionen für die mit der SWD und den beiden Fachthesauri erschlossenen Datenbanken zeigen weniger Übereinstimmung. In der Anwendung der SWD gilt zum einen das Prinzip des engen Schlagworts und zum anderen der Grundsatz, dass nur der Kern eines Dokuments, aber keine Randfragen berücksichtigt werden. Das heißt konkret, dass Dokumente, die mit der SWD erschlossen sind, in der Regel wenige und recht spezifische Schlagwörter enthalten. Die Indexierung in den Datenbanken von IZ und FIS Bildung dagegen deckt in der Regel alle relevanten Haupt- und Nebenaspekte eines Dokuments ab, sodass die Zahl der pro Dokument vergebenen Schlagwörter im Durchschnitt wesentlich größer ist als in dem SWD-erschlossenen Bestand der DDB.

b) Fachgebiete

Entsprechend dem Projektantrag waren für die Verknüpfung die Fachgebiete Soziologie, Methoden der Sozialforschung, Politikwissenschaft sowie Bildungsforschung zu bearbeiten. Die entsprechenden Bereiche wurden anhand der Systematiken der Thesauri ausgewählt.

Thesauri	ausgewählte Systematikgruppen	Anzahl der Terme*
SWD	9.2 Sozialwissenschaften allg., Soziologische Theorien, Statistik in den Sozialwissenschaften	562
	9.3 Sozialstruktur, Soziales Leben, Bevölkerung	2.016
	9.4 Arbeit, Arbeitswelt, Gewerkschaften	1.018
	9.5 Sozialpolitik, Sozialarbeit	635
	8.1 (Politik, Militär) Allgemeines, Politische Theorie	1.002
	8.2 Innenpolitik, Parteien	469
	8.3 Außenpolitik	203
	Summe	5.905
TheSoz	0.1 Wissenschaftstheorie, Methodologie, Methoden	265
	0.2.07 Theorien und Ansätze der Soziologie/ Sozialpsychologie	41
	0.3.07 wiss. Disziplinen und Teilbereiche der Soziologie/ Sozialpsychologie	66
	1. Gesellschaft	382
	2. Soziales Verhalten, Grundlagen und Erscheinungsformen	845
	3.1 Sozialwesen und Sozialpraxis	213
	3.3 Arbeit, Arbeitsmarkt (Auswahl)	198
	4.1 Erziehung, Bildung (Auswahl)	164
	4.2 Kommunikation, Sprache, Publizistik (Auswahl)	139
	4.3 Bevölkerung, Demographie	103
	4.4 Politik	1.032
	4.6.02 Wirtschaftspolitik	85
	Summe	3.533
ThesBild	Bildungsforschung	232
	Bildungssoziologie	150
	Pädagogische Kernbegriffe	500
	Summe	882

* incl. Dubletten wegen der Zuordnung zu mehreren Sachgruppen

Aufgrund der unterschiedlichen inhaltlichen Struktur von Universal- und Fachthesauri sind die Bereiche nicht deckungsgleich, so dass im Zuge der Bearbeitung auch Teilgebiete aus anderen Systematikgruppen herangezogen werden mußten. So enthält beispielsweise die SWD-Systematikgruppe „9 Soziologie, Gesellschaft...“ auch Begriffe zu „Bevölkerung“, die im TheSoz in einer eigenen Gruppe „Bevölkerung, Demographie“ erscheinen. Auf der anderen Seite enthält beispielsweise die Gruppe „0.1 Wissenschaftstheorie, Methodologie, Methoden“ des TheSoz auch Begriffe aus dem statistischen Bereich, der in der SWD der Systematikgruppe „29 Statistik“ zugeordnet ist.

Ausgangsbasis für die Erarbeitung von Crosskonkordanzen waren beim TheSoz ca. 3.400 Deskriptoren. In der SWD ergab sich für die Crosskonkordanz eine Menge von ca. 5.800 Deskriptoren. Diese Zahl hat sich gegenüber der im Zwischenbericht genannten Zahl verringert, weil zum einen Redundanzen im Laufe der Bearbeitung ausgefiltert wurden (manche Deskriptoren kommen in mehreren Systematikgruppen vor, wurden aber nur einmal bearbeitet). Zum anderen hat sich bei der Bearbeitung der Systematikgruppe 8.4 der SWD „Militär“ herausgestellt, dass sich diese Gruppe nicht gut zu einer Konkordanz mit dem TheSoz eignet. Für den Bereich „Militär“ deckt der TheSoz nur den sozialwissenschaftlichen Bereich ab, während die SWD in dieser Gruppe rund 2.300 z.T. recht detaillierte Deskriptoren und zahlreiche Eigennamen, z.B. von Waffen, verzeichnet. In ca. 90% der Fälle wären nur Verknüpfungen zu sehr entfernten Oberbegriffen im TheSoz möglich gewesen, die aber in der Recherche zu unsinnigen Treffermengen geführt hätten. Deshalb haben die Bearbeiterinnen entschieden, diese 2.300 SWD-Deskriptoren aus der Konkordanzarbeit auszuschließen. Aus dem Schlagwortmaterial des FIS Bildung (ThesBild) wurden anhand der Fachsystematik Bildung 882 Schlagwörter aus den Bereichen Bildungsforschung, Bildungssoziologie sowie sachgruppenübergreifende pädagogische Kernbegriffe ausgewählt und mit den entsprechenden Schlagwörtern der SWD verknüpft.

c) Typen von Relationen

In Tests vorab erprobt, hat sich auch während der Bearbeitung die Verwendung der folgenden Typen von Relationen als sinnvoll erwiesen:

- Äquivalenzen („=“)
 - identische Benennungen für identische Sachverhalte
 - verschiedene Benennungen für identische Sachverhalte
 - verschiedene Benennungen für ähnliche Sachverhalte, die in einem der Thesauri gleichgesetzt sind
- Oberbegriffe („<“)
- Unterbegriffe („>“)
- Verwandte Begriffe („><“)

Bei diesen Arten der Relationen kann die Verknüpfung jeweils mit einem einzelnen Term des anderen Thesaurus oder mit einer Verknüpfung von mehreren Termen erfolgen. Die 1:n-Relation kann eine logische UND- („+“) oder eine logische ODER-Verknüpfung („o“) beinhalten.

Die zunächst vorgesehene umgekehrte n:1-Relation wurde nicht hergestellt, da Schlagwortkombinationen in den Thesauri nicht als Deskriptoren behandelt sind. Eine solche n:1-Darstellung ist zur Bildung von Suchlogiken für verschieden erschlossene Datenbanken auch nicht erforderlich, da die Recherche immer mit einem oder mehreren Deskriptoren des Ausgangsthesaurus (Eigenthesaurus) beginnt, zu jedem einzelnen Deskriptor anschließend die Konkordanz im Zielthesaurus (Fremdthesaurus) ermittelt und auf diese Weise die neue Suchlogik sukzessive aufgebaut wird. Für Deskriptoren ohne sinnvolle Entsprechung wurde eine 0-Relation festgelegt. Dies ist dann der Fall, wenn einer der Thesauri beispielsweise Eigennamen für psychologische Tests, für Preise u.ä. oder aber eine sehr spezielle Benennung enthält, die nur mit einem sehr weiten Oberbegriff im anderen Thesaurus korrespondiert, und die Suche mit der Crosskonkordanz zu Treffermengen mit sehr viel Ballast führen würde.

d) Relevanz

Das Maß für die Übereinstimmung der verknüpften Deskriptoren wurde als hohe, mittlere oder geringe Relevanz definiert. Es wurde aufgrund der Annahme / Erfahrung eingeschätzt, wie passend das Retrievalergebnis sein wird, wenn von einem Suchbegriff in der ersten Datenbank auf den verknüpften Suchbegriff in der anderen Datenbank umgestiegen wird. Das Verhältnis von Recall und Precision wurde primär anhand der Definitionen der Benennungen bestimmt, nur in Einzelfällen empirisch anhand von Datenbankrecherchen. Es wurden Richtlinien für die Relevanzbeurteilung entwickelt, bei denen folgende Kriterien eine Rolle spielen:

- Vorzugsbenennungen im jeweils anderen Thesaurus
- Abstand in der Hierarchie bei hierarchischen Beziehungen
- Wahrscheinlichkeit von Verknüpfungen mit anderer Bedeutung bei Kombinationsrelationen.

1.4.2 Erarbeitung von Crosskonkordanzen

Da das Software-Tool SIS-TMS zur Erstellung, Pflege und grafischen Darstellung von Konkordanzen erst sehr spät in der Projektlaufzeit zur Verfügung stand und sich außerdem das Eingabetool als bedienungsunfreundlich und nachbesserungsbedürftig herausgestellt hat, wurden die Konkordanzen in Word- und später in Excel-Tabellen erstellt.

Die Bearbeitungsrichtung wurde zwischen DDB und IZ so aufgeteilt, dass die Konkordanz überwiegend ausgehend vom fremden in den eigenen Thesaurus erstellt wurde. Es wurden 5.836 Sachschlagwörter aus der SWD und 3.400 Deskriptoren aus dem TheSoz bearbeitet. Die Konkordanzen wurden nach der Erstellung von der jeweils anderen Partnerin überprüft und Problemfälle in 3-4 Arbeitstreffen pro Jahr besprochen. Es hat sich herausgestellt, dass die Verknüpfungen in den allermeisten Fällen reziprok sind. Daraus ergibt sich als Empfehlung für künftige Arbeiten, die Erstellung von Crosskonkordanzen zu teilautomatisieren: Die Verknüpfung wird in einer Richtung angelegt und bei Erarbeitung der Gegenrichtung bietet das Konkordanz-Management-System die reziproke Verknüpfung als Vorschlag. Dieses Verfahren spart eine erhebliche Menge an Schreibarbeit, lässt aber dem Bearbeiter die Kontrolle über die Verknüpfungen.

Die aus dem ThesBild ausgewählten 882 Deskriptoren wurden in einem ersten Arbeitsschritt automatisch mit den SWD-Schlagwörtern abgeglichen, wobei sich ein hoher Anteil von rund 80% zeichenidentischer Terme ergab. Diese wurden auf ihre tatsächliche Äquivalenz überprüft und die restlichen 20% der Schlagwörter intellektuell verknüpft.

Je nach Bearbeitungsrichtung ergaben sich unterschiedliche Anteile der Relationstypen und Relevanzbeurteilungen:

	SWD→TheSoz	TheSoz→SWD	ThesBild→SWD
Äquivalenz (=)	49,8%	86,5%	89,6%
hohe Relevanz	39,8%	81,9%	88,2
mittlere Relevanz	9,1%	4,5%	1,3
geringe Relevanz	0,9%	0,1%	0,1
Oberbegriff (<)	30,0%	4,9%	5,0%
hohe Relevanz	1,4%	0,1%	0,7
mittlere Relevanz	18,3%	3,0%	3,7
geringe Relevanz	10,3 %	1,8%	0,6
Unterbegriff (>)	1,3%	2,4%	1,2%
hohe Relevanz	--	0,5%	0,1
mittlere Relevanz	0,9%	1,9%	1,0
geringe Relevanz	0,4%	--	0,1
Verwandter Begriff (><)	8,7%	3,1%	3,1%
hohe Relevanz	0,7%	0,6%	0,1
mittlere Relevanz	5,0%	1,7%	2,3
geringe Relevanz	3,0%	0,8%	0,7
keine Relation möglich (0)	10,2%	3,1%	1,1%
Basis N=	5.836	3.400	882

(Auswertung der Eingabetabellen mit SPSS. Die geringfügigen Abweichungen der Werte für ThesBild von den Zahlen im Zwischenbericht beruhen auf der Überarbeitung in Einzelfällen)

Es ist festzustellen, dass sich für diese Gebiete überwiegend äquivalente Beziehungen herstellen ließen. Von der SWD ausgehend sind mehr Oberbegriffs- und 0-Relationen zu verzeichnen. Von den Fachthesauri ausgehend ist ein hohes Maß an Term-Äquivalenzen mit der SWD vorhanden. Dabei hat die Auswahl von fachlichen Kernbegriffen für die Konkordanz von ThesBild mit der SWD vermutlich auch einen Einfluss auf den größeren Anteil von äquivalenten und hochrelevanten Verknüpfungen. Die Differenzen ergeben sich aus dem unterschiedlichen Maß an Prä- und Postkoordination in den Thesauri sowie aus der größeren Hierarchietiefe und dem höheren Bestand an Eigennamen in der SWD gegenüber dem TheSoz.

Verknüpfungen mit hoher Relevanz ergaben sich erwartungsgemäß vor allem bei den Termen mit Äquivalenzbeziehung, während die hierarchischen Konkordanzrelationen überwiegend mit mittlerer Relevanz eingestuft wurden. Gering relevante Verknüpfungen haben einen relativ kleinen Anteil, was im Hinblick auf sinnvolle Rechercheergebnisse mit diesen Konkordanzrelationen auch plausibel ist.

Die erstellten Konkordanzen sind komplett in SIS-TMS gespeichert und können dort abgerufen werden. Sie liegen außerdem als Excel-Tabellen vor.

1.4.3 Exemplarische Tests der Crosskonkordanzen anhand von Datenbanken

Vorgehensweise

Zur stichprobenhaften Überprüfung wurden insgesamt 130 der erarbeiteten Crosskonkordanzen zu Datenbankrecherchen genutzt. Sie wurden so ausgewählt, dass alle Relationsarten in Kombination mit allen vorkommenden Relevanzeinschätzungen getestet werden konnten. Da das Transfertool von AP 11 zu diesem Zeitpunkt noch nicht zur Verfügung stand, wurde der Rechercheablauf "per Hand" simuliert. Ausgewertet wurden die Rechercheergebnisse mit

"TheSoz→SWD" bzw. "ThesBild→SWD" konkordanten Deskriptoren im Web-OPAC DDB⁹ bzw. im Verbund-OPAC des Bayerischen Bibliotheksverbundes (BVB) sowie mit "SWD→TheSoz" konkordanten Deskriptoren in der Datenbank SOLIS ab 1990 (CD-ROM). Die gefundenen Titel wurden jeweils bzgl. hoher, mittlerer und geringer Relevanz bewertet, wobei dieser Einschätzung folgende Maßstäbe zugrunde lagen:

- hohe Relevanz: der gesuchte Sachverhalt ist der oder ein wesentlicher Hauptaspekt im gefundenen Dokument
- mittlere Relevanz: der gesuchte Sachverhalt ist ein nachrangiger, aber nicht marginaler Aspekt im gefundenen Dokument
- geringe Relevanz: der gesuchte Sachverhalt ist kein oder nur ein marginaler Aspekt im gefundenen Dokument

Bei großen Treffermengen wurden nur 30 Treffer ausgewertet.

Ergebnisse

Legende zu den Tabellen:

=	Äquivalenz	h	Treffer von hoher Relevanz
<	Oberbegriff	m	Treffer von mittlerer Relevanz
>	Unterbegriff	g	Treffer von geringer Relevanz
><	Verwandter Begriff		

a) TheSoz (SOLIS) → SWD (OPACs DDB/BVB)

Ausgehend von einem TheSoz-Dekriptor wurde mit dem/den verknüpften Schlagwort/Schlagwörtern der SWD gesucht. 27 Konkordanzrelationen wurden getestet und 475 Treffer bewertet.

Angaben in Prozent

	DDB			BVB			Mittelwert aus DDB und BVB		
Rel	h	m	g	h	m	g	h	m	g
=	70	26	4	73	20	7	71,5	23	5,5
<	24	21	55	32	23	45	28	22	50
>	58	34	8	49	27	24	53,5	30,5	16
><	42	43	15	36	41	23	39	42	19

Lesehilfe: Bei den mit Äquivalenzrelation (=) verknüpften Deskriptoren lieferte eine Suche im DDB-OPAC 70% hochrelevante Treffer, 26% mittelrelevante Treffer und 4% geringrelevante Treffer. Die Relevanz der Treffer im BVB-OPAC war zu 73% hoch, zu 20% mittel und zu 7% gering.

⁹ Im DDB-OPAC wurde das jeweilige Schlagwort mit Phrasensuche gesucht, diese Einschränkung war im BVB-OPAC nicht nötig. Zwischenzeitlich wurde eine neue Version des DDB-OPACs eingeführt. Die Unterscheidung zwischen Schlagwörtern vor und nach 1986 sowie die Phrasensuche fallen weg. Weitere Neuerungen sind die Begrenzung auf den Frankfurter Bestand sowie der Wegfall der Mengenbegrenzung der Trefferanzeige. Dadurch kann sich der Recall verändern, so dass die Test-Ergebnisse in dieser Hinsicht u.U. anders zu bewerten sind.

b) ThesBild (FIS Bildung) → SWD (OPACs DDB/BVB)

Ausgehend von einem ThesBild-Deskriptor wurde mit dem/ den verküpften Schlagwort/ Schlagwörtern der SWD gesucht. 17 Konkordanzrelationen wurden getestet und 342 Treffer bewertet.

Angaben in Prozent

	DDB			BVB			Mittelwert aus DDB und BVB		
Rel	h	m	g	h	m	g	h	m	g
=	71	20	9	68	24	8	69,5	22,0	8,5
<	67	30	3	60	40	0	63,5	35,0	1,5
>	31	53	16	31	58	11	31,0	55,5	13,5
>>	83	17	0	100	0	0	91,5	8,5	0

c) Test SWD → TheSoz (SOLIS)

Ausgehend von einem SWD-Schlagwort wurde mit dem/ den verknüpften TheSoz-Deskriptor/en gesucht. 86 Konkordanzrelationen wurden getestet und 1217 Treffer bewertet.

Angaben in Prozent

	SOLIS		
Rel	h	m	g
=	75	17	8
<	32	23	44
>	67	24	9
>>	37	13	50

Bei den Äquivalenzbeziehungen liegt der Anteil der hoch- und mittelrelevanten Titel über 90%, das heißt, dass der gesuchte Sachverhalt entweder der zentrale oder ein wesentlicher Aspekt des in der Zieldatenbank gefundenen Dokuments ist.

Bei den Oberbegriffsrelationen streuen die Ergebnisse beim Umstieg von SWD zu TheSoz bzw. umgekehrt stark in Abhängigkeit davon, wie weit der Ober- vom Unterbegriff entfernt ist. Es gibt zwar Fälle mit einem sehr hohen Anteil an hochrelevanten Titeln, doch liefern diese Verknüpfungen auch etwa zur Hälfte geringrelevante Titel. Allerdings gibt es bei den Oberbegriffsrelationen auch „Ausreißer“, deren Ergebnis sehr schlecht ist. Hier wäre zu überlegen, ob man Relationen zu sehr weiten Oberbegriffen nicht besser auf null setzt, statt einen sehr hohen Recall und eine sehr geringe Precision zu riskieren. Für die Verbindung ThesBild → SWD sieht das Ergebnis sehr viel besser aus, hier liefern die Oberbegriffsrelationen zu 98% hoch- und mittelrelevante Treffer.

Die Unterbegriffsrelationen liefern mit 84-91% hoch- und mittelrelevanten Treffern gute Ergebnisse.

Die Verknüpfungsart Verwandter Begriff streut im Ergebnis stark. Die Konkordanzen von den Fachthesauri zur SWD liefern mit 81-100% hoch- und mittelrelevanten Treffern gute Ergebnisse, während beim Umstieg von der SWD zum Fachthesaurus die gering relevanten Treffer bei 50% liegen. In dieser Gruppe treten „Ausreißer“ allerdings häufiger auf als in den anderen Gruppen.

Der Anteil der hoch- bzw. gering-relevanten Treffer der Rechercheergebnisse entspricht in den ganz überwiegenden Fällen den Relevanzeinschätzungen der Konkordanz. Wo dies im

Laufe künftiger praktischer Erfahrungen nicht der Fall sein sollte, müsste überlegt werden, ob die Relevanzeinschätzungen der Konkordanz zu korrigieren sind oder ob besondere Bedingungen der Datenbank (z.B. fachliches Spektrum) das Ergebnis beeinflussen.

Insgesamt zeigen die Testergebnisse, dass das konzeptionell-methodische Verfahren zur Erstellung von Crosskonkordanzen tragfähig ist, und die bisherige inhaltliche Erarbeitung von Konkordanzrelationen und Relevanzbeurteilungen fach- und datenbankübergreifende Suchen in diesen Gebieten erleichtert.

Empfehlungen zur Optimierung des erzielten Ergebnisses und für weitere Konkordanzarbeiten

Die Konkordanzen sollten durchgehend reziprok aufgebaut werden, d.h. auch bei Termen, bei denen zur Konkordanzbildung im Zielthesaurus auf "fachfremde" Systematikgruppen zugegriffen wurde. Nicht-äquivalente Verknüpfungen sollten systematisch an wichtigen Datenbanken überprüft werden, um die Qualität der Suchergebnisse zu erhöhen. Zu achten wäre hierbei insbesondere auf

- zu weit entfernte Ober- und verwandte Begriffe, deren Verknüpfungen auf null zu setzen sind
- Ergänzung von ODER-Verknüpfungen aufseiten der SWD in der Konkordanz zum ThesBild
- UND-Kombinationen, sofern sie das Suchergebnis zu stark einschränken, hier wäre die Beschränkung auf ein Schlagwort als Oberbegriff sinnvoller.

Bei weiteren Konkordanzarbeiten sollten die während der empirischen Arbeit entstehenden Regeln und Kriterien dokumentiert und systematisiert werden. Hierbei sollten außerdem Entscheidungshilfen für die Behandlung von Randgebieten formuliert werden (wie im vorliegenden Falle zum Bereich „Militär“).

Die Einbeziehung von Geographika und wichtigen Körperschaften in die Konkordanzen hat sich in den Tests als notwendig herausgestellt und sollte von den Spezialthesauri zur SWD erstellt werden.

1.5 Analyse der Übernahmebedingungen der Crosskonkordanzen in die Dienstleistungen der DDB (UAP 14)

Crosskonkordanz Realeinsatz-Szenario (SWD, TheSoz, ThesBild)

Die Crosskonkordanzen zwischen den Bereichen Sozialwissenschaften und Politik der SWD und ausgewählten Gebieten des Thesaurus Sozialwissenschaften sowie zwischen ausgewähltem Schlagwortmaterial des FIS Bildung aus den Bereichen Bildungsforschung und Bildungssoziologie und der SWD sollen der inhaltlichen Suche in heterogenen Datenbeständen dienen. Durch den Einsatz von Crosskonkordanzen kann der Benutzer seine inhaltliche Suche im Bereich der Sozialwissenschaften, der Politik und der Bildungsforschung über mehrere Datenbestände ausweiten, die mit einem der drei Thesauri erschlossen sind, ohne alle Thesauri kennen und das ihm vertraute System verlassen zu müssen. Folgende Szenarien sind denkbar:

a) DDB-OPAC -> SOLIS / FIS Bildung

Weiterleitung der Suchanfrage von einer bekannten Datenbank zu einer weiteren, ausgewählten Datenbank

- Aufruf des DDB OPAC
- inhaltliche Suche mit einem oder mehreren Schlagwörtern
- Anzeige der Kurz-Trefferliste des DDB-OPAC
- falls in der Suche ein Schlagwort enthalten ist, das mit dem Kennzeichen „Konkordanzeintrag erstellt“ versehen ist, wird vom System ein Button angeboten, mit welchem man die erweiterte Suche aktivieren kann: „Möchten Sie Ihre Suche auch in der Datenbank SOLIS / FIS Bildung durchführen? Bitte hier klicken“
- jedes SWD-Schlagwort wird über das Transfertools zu SIS-TMS weitergeleitet, der konkordante Term ermittelt und über das Transfertools in der vorhandenen Suchlogik ersetzt
- die neu generierte Suchanfrage ist dann z.B. an SOLIS im Webangebot der GBI oder an die Literaturdatenbank des FIS Bildung weiterzuleiten
- die jeweilige Datenbank liefert in einem neuen Fenster die Titelliste der Treffer kostenlos
- weitere Informationen, z.B. Autor/in, bibliographische Angaben, Abstract und ggf. auch der Volltext können gegen Entgelt direkt nachgefordert werden.

b) DDB-OPAC -> Virtuelle Bibliothek Sozialwissenschaften (ViBSoz) / Informationsverbund Bildung - Sozialwissenschaften - Psychologie (Informationsverbund BSP)

Weiterleitung der Suchanfrage von einer bekannten Datenbank zu einem Verbund von Datenbanken

- Aufruf des DDB OPAC
- inhaltliche Suche mit einem oder mehreren Schlagwörtern
- Anzeige der Kurz-Trefferliste des DDB-OPAC
- falls in der Suche ein Schlagwort enthalten ist, das mit dem Kennzeichen „Konkordanzeintrag erstellt“ versehen ist, wird vom System ein Button angeboten, mit welchem man die erweiterte Suche aktivieren kann: „Möchten Sie Ihre Suche auch in der Datenbank SOLIS / FIS Bildung durchführen? Bitte hier klicken“
- die SWD-Suchanfrage wird an ViBSoz bzw. den Informationsverbund BSP weitergeleitet
- ViBSoz bzw. der Informationsverbund BSP liefern über ihre Transfermodule zu jeder im Verbund angebotenen Datenbank den jeweils passenden Transfer von Termen zur Suche
- die Suchergebnisse aller angesprochenen Datenbanken werden als Titelliste kostenlos auf der Benutzeroberfläche angezeigt
- weitere Informationen, z.B. Autor/in, bibliographische Angaben, Abstract und ggf. auch der Volltext können gegen Entgelt direkt nachgefordert werden.

Pflege und Update der Crosskonkordanz

In den im Projekt bearbeiteten Bereichen der Sozialwissenschaften und Politik fallen im Laufe eines Jahres rund 2.000 Neuansetzungen, Änderungen und Löschungen in der SWD an, was bedeutet, dass durchschnittlich 40 Änderungen pro Woche vom Fachreferenten auch in der Konkordanz bearbeitet werden müssten. Ein Update der Crosskonkordanz ist bei jeder Veränderung, auch bei solchen in den Relationen (synonyme Verweisungen, Oberbegriffe, Unterbegriffe, Splits, Umlenkungen etc.), vorzunehmen, und zwar grundsätzlich „per Hand“, da die Verknüpfungen zu den anderen Thesauri bei allen terminologischen Änderungen intellektuell überprüft und gegebenenfalls überarbeitet werden müssen. Aus Sicht DDB ist ein in den SWD-Arbeitsablauf integrierter Geschäftsgang für die Pflege der Crosskonkordanz

notwendig, in welchem es erforderlich ist, bei jeder Veränderung in der SWD in den Bereichen Sozialwissenschaften und Politik die Crosskonkordanz zu überprüfen und ggf. zu verändern. Die SWD wird zur Zeit in einem kooperativen Verfahren von mehreren hundert Bibliotheken im ILTIS Der Deutschen Bibliothek online und mit der ständigen Option, Änderungen, Ergänzungen und Löschungen vornehmen zu können, geführt.

Der Einsatz von SIS-TMS für die Konkordanzpflege in einem integrierten Verfahren kann aus den folgenden Gründen nicht in Betracht gezogen werden:

- Ein wesentlicher Nachteil ist, dass die beteiligten Thesauri komplett in das Tool integriert werden müssen. Das bedeutet, dass für die SWD eine doppelte Datenhaltung sowohl in SIS-TMS als auch in der jetzigen Umgebung notwendig wäre.
- Die Erfassungskomponente Entry Form von SIS-TMS hat sich im Projekt als schwerfällig und sehr bedienungsunfreundlich erwiesen. Im Projekt wurde die Crosskonkordanz aus diesem Grunde in Excel erfasst. Zwar wurde vom IZ Sozialwissenschaften ein Vorschlag zur Verbesserung des Eingabetools erstellt, den ICS-FORTH außerhalb von CARMEN umsetzen wird, jedoch liegt DDB zur Zeit noch kein testfähiges Modul vor.
- SIS-TMS kann in die ILTIS-Katalogisierungsumgebung nicht integriert werden, da wie oben erwähnt die Konkordanzpflege immer intellektuell und in SIS-TMS erfolgen muss. Die SWD-Bearbeiter müssten also auf Dauer eine zusätzliche Arbeitsumgebung bedienen.
- Es ist unsicher, ob SIS-TMS die Größenordnung an parallelen Zugriffen wie sie für die Arbeitsumgebung der SWD notwendig ist, ermöglicht.
- Der Einsatz von SIS-TMS ist mit hohen Risiken behaftet:
 - Ein Tool für ein integriertes Verfahren muss von hoher Zuverlässigkeit und Verfügbarkeit sein. Von SIS-TMS gibt es keine vergleichbare Referenzinstallation.
 - Bei ICS-FORTH handelt es sich um eine kleine Forschungsgruppe. Dauerhafte Pflege und Support sind fraglich.

DDB plant, für die Crosskonkordanz ein noch näher zu bestimmendes, auf die Bedürfnisse von DDB besser zugeschnittenes und auch für andere Thesaurus-Konkordanzen geeignetes Pflege-Tool einzusetzen. Die jeweils aktuelle Konkordanz könnte dann in größeren Abständen offline an das IZ Sozialwissenschaften geliefert werden, falls das IZ Sozialwissenschaften SIS-TMS weiterhin zur Haltung der Crosskonkordanz verwendet. Ebenso könnten die Änderungen in der SWD in größeren Abständen offline oder in großen Abständen eine komplette Datenneulieferung geliefert werden. Die SWD-ID-Nummer müsste allerdings als zusätzliches Attribut zur SWD-Ansetzung mitgeführt werden, da im Änderungsdienst ein Austausch von Ansetzungsänderungen über die SWD-ID-Nr. erfolgt. Der „Altstand der Ansetzung“ wird hingegen nicht mitgeführt.

1.6 Verbindung zwischen den Crosskonkordanzen Thesauri und Klassifikationen (UAP 10)

Für die qualitative Analyse der Verbindung zwischen den Crosskonkordanzen wurden drei Szenarien ausgewählt, und zwar der Zusammenhang zwischen

- a) IZ-Klassifikation - TheSoz (alphabetisch) - SWD
- b) TheSoz (alphabetisch) - IZ-Klassifikation - BK/ RVK
- c) BK/ RVK - IZ-Klassifikation -TheSoz (systematisch)

Ziel der Analyse ist die Prüfung der Ausgangshypothese, dass es möglich sein müsste, mit Hilfe der Crosskonkordanzen von einem nur durch eine Klassifikation (eine Bibliothekssystematik) erschlossenen Datenbestand zu einem nur durch einen Thesaurus erschlossenen Datenbestand - oder umgekehrt - zu wechseln, ohne sich vorher das jeweils andere System aneignen zu müssen.

Für a) und b) wird an Beispielen jeweils der Zusammenhang zwischen Klassifikation und Thesaurusdeskriptoren empirisch aus der Literaturlatenbank des IZ abgeleitet und mit Hilfe der Crosskonkordanz eine Recherche in einer anderen Datenbanken durchgeführt. Unter c) wird der Versuch unternommen, eine retrievalmäßige Verknüpfung zwischen einem systematischen Thesaurus einerseits - unter Benutzung der dort vorhandenen klassifikatorischen Elemente - und einer Klassifikation andererseits herzustellen und an einem bibliothekarisch/ dokumentarischen Datenbestand zu testen.

zu a) IZ-Klassifikation - TheSoz (alphabetisch) - SWD

Für 6 Klassen der IZ-Klassifikation wurden aus SOLIS, der Literaturlatenbank des IZ, die jeweils mit den Dokumenten verknüpften Deskriptoren ermittelt.

Wie unter 1.4.1.a beschrieben, haben die Indexierungsregeln und -konventionen des IZ relativ viele Deskriptoren pro Dokument zur Folge. Das erklärt die hohe Anzahl von Deskriptoren (Token), die insgesamt in Dokumenten mit einer bestimmten Klassifikation auftreten. Die einzelnen Deskriptoren (Types) sind sehr unterschiedlich besetzt, die breite Streuung entsteht vor allem durch eine große Anzahl von sehr gering besetzten Deskriptoren. Um dies in der Tabelle darzustellen, wurden die am häufigsten vorkommenden Deskriptoren ermittelt, die jeweils in mindestens 30% bzw. 10% der Dokumente vorkommen. In Dokumenten mit der Klassifikation „Familiensoziologie“ sind es z.B. nur 2 bzw. 14 von 797 Deskriptoren, die in mindestens 30% bzw. 10% der Dokumente vorkommen. „Familie“ ist in 57,2% und „Bundesrepublik Deutschland“ in 36,6% der Dokumente enthalten (s.u.). Mehr als die Hälfte der Deskriptoren (497) ist nur in 1 oder 2 Dokumenten indexiert.

IZ-Klassifikation – Deskriptoren

IZ-Klassifikation	Langtext	Treffer*	Deskriptoren (Token)	Deskriptoren (Types)	Types in mind. 30% der Treffer	Types in mind. 10% d. Treffer
10209	Familiensoziologie	355	3686	797	2	14
10212	Agrarsoziologie	132	1484	589	2	17
10214	Kriminal-, Rechtssoziologie	294	3193	922	2	9
10215	Medizinsoziologie	227	2158	729	1	9
10504	Politische Soziologie	493	5155	1208	1	5
10803	Interpersonelle Kommunikation	136	1536	631	2	11

*es wurde nur mit der Hauptklassifikation recherchiert. In jedem Dokument ist dies eine der vergebenen Klassifikationen, der das Dokument schwerpunktmäßig zugeordnet ist.

Beispiel:

Deskriptoren zur Klassifikation 10209 (Familiensoziologie)

1. 000203 FAMILIE
 2. 000130 BUNDESREPUBLIK DEUTSCHLAND
 3. 000075 EHE
 4. 000060 PARTNERSCHAFT
 5. 000059 SOZIALE BEZIEHUNGEN
 6. 000056 KIND
 7. 000048 SOZIALER WANDEL
 8. 000047 EHESCHIEDUNG
 9. 000045 FRAU
 10. 000043 GENERATIONENVERHAELTNIS
 11. 000041 HEIRAT
 12. 000040 ELTERN-KIND-BEZIEHUNG
 13. 000038 HISTORISCHE ENTWICKLUNG
 14. 000037 LEBENSWEISE
- usw.

Werden die häufigsten Deskriptoren über die Crosskonkordanz mit der SWD verknüpft und die betreffenden Schlagwörter für eine Suche im DDB-OPAC verwendet, führt das zu folgendem Ergebnis:

IZ-Klassifikation – Deskriptoren – SWD

IZ-Klassifikation	Häufigstes Schlagwort nach Transfer	Treffer im DDB-OPAC	Relevante Treffer*	Transfer von 2 Schlagwörtern **	Treffer im DDB-OPAC (ODER-Verknüpfung)	Relevante Treffer*
10209	Familie	2889	20%	Familie Ehe *** Lebensgemeinschaft	3715	23,3%
10212	Landwirtschaft	836	23,3%	Landwirtschaft Ländlicher Raum	1288	30,0%
10214	Kriminalität	290	56,7%	Kriminalität Strafrecht	1392	10,0%
10215	Gesundheit	796	30,0%	Gesundheit Krankheit	2551	13,3%
10504	Politische Kultur	378	66,7%	Politische Kultur Demokratie	980	63,3%
10803	Kommunikation	748	43,3%	Kommunikation Interaktion	1040	50,0%

* es wurden jeweils die ersten 30 Treffer ausgewertet

** geographische Schlagwörter wurden nicht berücksichtigt

*** von dem IZ-Deskriptor „Ehe“ aus besteht eine Konkordanzrelation sowohl mit dem SWD-Schlagwort „Ehe“ als auch mit „Lebensgemeinschaft“

Bei den am häufigsten vorkommenden Schlagwörtern, die also für eine bestimmte Klassifikation besonders charakteristisch sind, liegt die Relevanz des Rechercheergebnisses zwischen 10 und 66%. Geographische Schlagwörter wurden dabei ausgeschlossen, weil sie sachlich völlig unspezifisch sind und, besonders im Fall von „Bundesrepublik Deutschland“, extrem große Treffermengen bringen. Als relevant wurden Dokumente eingeschätzt, die entsprechend der Ausgangsfragestellung einer bestimmten Klassifikation zuzuordnen sind. Bemerkenswert ist, dass eine Erweiterung der Recherche durch das zweithäufigste Schlagwort die Ergebnisse nur in der Hälfte der Fälle etwas verbesserte.

Tendenziell ist festzustellen, dass bei denjenigen Schlagwörtern die Relevanz des Retrievalergebnisses höher ist, die semantisch der soziologischen Terminologie näher stehen bzw. der soziologischen Klassifikation, von der der Transfer ausging, z.B. „Politische Kultur“ oder „Kommunikation“. Das Umgekehrte gilt für „Landwirtschaft“, „Strafrecht“ oder

„Demokratie“, die der (Familien-)Soziologie ferner stehen und das Retrievalergebnis eher verschlechtern.

zu b) TheSoz (alphabetisch) - IZ-Klassifikation – BK/RVK

Für 5 Deskriptoren des TheSoz wurden in SOLIS, der Literaturdatenbank des IZ, die Hauptklassifikationen ermittelt, denen die entsprechenden Dokumente zugeordnet waren. Die Hauptklassifikationen pro Deskriptor sind breit gestreut, ihre Anzahl liegt zwischen 7 und 31. Wenige Klassifikationen kommen relativ häufig vor, d.h. die meisten Dokumente verteilen sich auf nur wenige Klassifikationen. Eine weitergehende Analyse wurde deshalb auf diejenigen Klassifikationen eingeschränkt, die den größeren Teil der Dokumente (mindestens 50 bzw. 75%) abdecken.

Deskriptor - Klassifikationen

Deskriptor	Treffer (h)	Haupt-Klassifikationen	h> 50% *	h> 75% *	RVK				BK			
					> 50% **	> 75% **	Treffer ***	Treffer mit Deskriptor	> 50% **	> 75% **	Treffer *****	Treffer mit Deskriptor
Adoption	109	21	3	7	18	43	2543	112	11	16	716	81
doppelte Staatsangehörigkeit	13	7	2	3	22	25	****		18	31	1140	2
Drogenkriminalität	44	13	2	3	15	20	15496	35	16	17	1421	13
Kleinfamilie	157	28	1	4	7	20	4029	1	10	27	2210	1
symbolische Politik	99	31	4	10 (=71%) 17 (=85%)	25	>200	****		7	21	1161	2

* d.h. mindestens 50% bzw. 75% der Treffer mit dem jeweiligen Deskriptor verteilen sich auf x Hauptklassifikationen

** Anzahl der Klassen, die mit den „>50%“ bzw. „>75%“-Klassifikationen verknüpft sind

*** Treffer einer Recherche im OPAC der UB Regensburg mit den „>50%“-Klassifikationen

**** Wegen der großen Zahl an Klassifikationen und der zu erwartenden Menge an Treffern wurden diese aufwändigen Recherchen nicht durchgeführt

***** Treffer einer Recherche mit den „>50%“ Klassifikationen in den USB-Beständen von ViBSoz (Virtuelle Fachbibliothek Sozialwissenschaften)

Am Beispiel „Adoption“ ist zu sehen, dass 109 Dokumente mit diesem Deskriptor 21 Hauptklassifikationen zugeordnet wurden, wobei sich jedoch mehr als die Hälfte der Dokumente (61) auf nur 3 Klassen verteilt, mehr als drei Viertel der Dokumente auf 7 Klassen. Die Crosskonkordanz der Klassifikationen mit den 3 bzw. 7 Klassen führt in der RVK zu 18 bzw. 43 Klassen, in der BK zu 11 bzw. 16 Klassen.

Eine exemplarische Suche mit den durch die Crosskonkordanz verknüpften 18 Klassifikationen der RVK im OPAC der UB Regensburg brachte 2543 Treffer. Ein analoges Vorgehen für die Deskriptoren „Drogenkriminalität“ und „Kleinfamilie“ führte zu 15496 bzw. 4029 Treffern. Das gleiche Verfahren wurde für eine exemplarische Suche mit der BK in den USB-Beständen von ViBSoz (Virtuelle Fachbibliothek Sozialwissenschaften) durchgeführt, wobei der Transfer von der IZ-Klassifikation zur BK genutzt wurde.

Ohne die Relevanz der Suchergebnisse im einzelnen zu prüfen, kann man bei diesen Treffermengen davon ausgehen, dass nur ein kleiner Teil für die Ausgangsfragestellung relevant ist.

Um dies gegenzuprüfen wurde direkt mit dem Deskriptor in der jeweiligen Ziel-Datenbank gesucht, das Ergebnis waren wesentlich geringere Trefferzahlen.

Bei der Anwendung der Crosskonkordanzen wurde für die genannten Fälle die Art der Relation zwischen den Klassifikationen (= / < / > / ><) und ihre Relevanz nicht genutzt.

Ergebnis von a) und b)

Anhand der exemplarischen Analysen konnte gezeigt werden, dass der Umstieg von der IZ-Klassifikation über TheSoz-Deskriptoren auf die SWD einen hohen Recall und relativ geringe Precision im Retrievalergebnis liefert. Akzeptable Precision kann jedoch erreicht werden bei Deskriptoren, die semantisch eindeutiger der Ausgangsklassifikation zuzurechnen sind. Der Weg von den Thesaurus-Deskriptoren über die IZ-Klassifikation zu RVK und BK führt zu Retrievalergebnissen mit noch höherem Recall und sehr niedriger Precision, d.h. zu viel Ballastinformation. Es liegt die Vermutung nahe, dass dieser methodische Ansatz für den Transfer weniger Entwicklungspotential bietet.

Zu untersuchen wäre noch für beide Fälle, welche Wirkung die Variablen „Art“ und „Relevanz der Konkordanzrelation“ sowie die Anzahl der für den Transfer verwendeten Deskriptoren/ Klassifikationen haben.

c) BK/ RVK – TheSoz(systematisch) – IZ-Klassifikation

Der im folgenden entwickelte und beispielhaft getestete Ansatz unterscheidet sich von den unter a) und b) dargelegten Verfahren dadurch, dass eine Verbindung zwischen einer Klassifikation und einem systematisch gegliederten Thesaurus nicht über die Klassifikations-Notation einerseits und einen Deskriptor andererseits hergestellt wurde, sondern dass die Thesaurus-Systematik dazu benutzt wurde, über die entsprechende Thesaurus-Notation eine ganze Gruppe von Thesaurusdeskriptoren mit einer Klassifikations-Notation zu verknüpfen. Hierzu hätte sich eine Verknüpfung zwischen RVK bzw. BK und dem IZ-Thesaurus angeboten. RVK, BK und IZ-Thesaurus werden jedoch zur inhaltlichen Erschließung sehr verschiedenartiger Datenbestände verwendet. Außerdem ist der Abgleich von Treffern in verschiedenen Datenbanken auch aus technischen Gründen sehr mühsam. Deshalb wurde der Entschluss gefasst, den Abgleich innerhalb ein- und desselben Datenbestandes vorzunehmen. Für diese Untersuchung wurde die IZ-Literaturdatenbank SOLIS ausgewählt. Weil SOLIS sowohl über eine Klassifikation als auch über einen systematischen Thesaurus verfügt, war es möglich, parallele Recherchen mit einer Klassifikations- und einer Thesaurus-Notation durchzuführen und das Ergebnis physikalisch abzugleichen, d.h. die Schnittmengen zu errechnen.

Da die Verknüpfungen zwischen der Regensburger Verbundklassifikation und der IZ-Klassifikation sowie zwischen der Basisklassifikation und der IZ-Klassifikation bereits vorhanden waren, konnte indirekt eine Brücke zwischen RVK und IZ-Thesaurus bzw. zwischen BK und IZ-Thesaurus geschlagen werden. Die jeweilige Verknüpfung zwischen RVK bzw. BK und IZ-Klassifikation wurde deshalb zu Vergleichszwecken an der linken Seite der Übersicht mit ausgewiesen. Außerdem wurden – wie aus der Übersicht ersichtlich – die zehn verschiedenen Verknüpfungsarten herangezogen. Auf diese Beispiele – die auch schon in den vorhergehenden Tests Klassifikation gegen Klassifikation verwendet wurden – wurde zurückgegriffen, obwohl sie teilweise für den Verwendungszweck nur bedingt geeignet sind. Immerhin wurde auf diese Weise nicht nur eine Positivauslese getroffen, sondern es konnte auch gezeigt werden, dass eine sinnvolle Verknüpfung (wie auch zwischen zwei Klassifikationen) nicht immer herzustellen ist.

Beim Vergleich Klassifikation/ Thesaurus wurde ähnlich vorgegangen wie beim Vergleich Klassifikation/ Klassifikation, d.h. die zu vergleichenden Paare wurden nach semantischen

Ähnlichkeiten ausgewählt. Dabei wurden aber auf Klassifikationsseite nur konkret zur Indexierung verwendete Klassen, d.h. die unterste Ebene, auf Thesaurus-Seite jedoch die dem einzelnen Deskriptor übergeordneten Notationen, welche Gruppen von Deskriptoren repräsentieren, herangezogen. Mit beiden Instrumenten wurde in SOLIS recherchiert und die Ergebnisse miteinander gekreuzt, um die Schnittmenge festzustellen. Anschließend war es möglich, die Qualität der Verknüpfung zu bewerten. Eine Übertragung der innerhalb SOLIS gewonnenen Erkenntnisse auf andere Datenbanken ist jedoch nur bedingt möglich, weil es sich bei SOLIS um eine auf einen bestimmten Fachbereich ausgerichtete Datenbank handelt und auch die inhaltlichen Erschließungsinstrumente speziell für diesen Fachbereich entwickelt wurden. Eine Konkordanz z.B. zwischen der RVK und dem IZ-Thesaurus und ein Vergleich der Treffer im Katalog einer Universalbibliothek mit den Treffern in SOLIS könnte daher u.U. zu ganz anderen Ergebnissen führen.

Da auch die der IZ-Klassifikation zugeordnete RVK- bzw. BK-Klasse angezeigt werden sollte, reichte der Platz auf den Übersichten nicht aus, um auch die absoluten Trefferzahlen auszuweisen, die zu den teilweise extremen Prozentzahlen führten (die Trefferzahlen waren auf der Klassifikationsseite meistens vierstellig, auf der Thesaurusseite vier- bis fünfstellig). Zunächst ist erstaunlich, dass die Verwendung einer Thesaurus-Notation, hinter der doch eine Zahl sehr ähnlicher Schlagwörter zu vermuten ist, meist zu einer noch größeren Zahl von Dokumenten führte, als das schon ziemlich grobe Instrument der Klassifikation. Es bleibt aber festzustellen, dass die Schnittmengen im allgemeinen recht groß waren, d.h. es fällt auf, dass ein recht großer Anteil der mit der Klassifikation gefundenen Dokumente auch in der Menge enthalten war, die mit der Thesaurus-Notation gefunden wurde. Andererseits übertraf die mit der Thesaurus-Notation gefundene Menge die (meist schon sehr große) Menge der mit der Klassifikation gefundenen Dokumente zum Teil noch erheblich. Anders ausgedrückt: Der Übergang von der Klassifikation zum Thesaurus führt zu großer Redundanz. Der umgekehrte Weg wäre meistens ergiebiger: nach der Suche mit einer Thesaurus-Notation würde der Übergang zur entsprechenden Klassifikations-Notation eine geringere, aber präzisere Treffermenge hervorbringen.

Fazit: Innerhalb SOLIS lassen sich Strukturen herausarbeiten, die eine Konkordanz zwischen Klassifikation und systematischem Thesaurus als nicht ganz unsinnig erscheinen lassen. Eine Übertragbarkeit der Ergebnisse ist aber nur mit Vorbehalt möglich.

Übersicht: Testverknüpfungen von der Regensburger Verbundklassifikation über die IZ-Klassifikation zur IZ-Thesaurus-Notation

Regensburger Verbund- klassifikation		Beziehung (Relevanz)	IZ Klassifikation		IZ-Thesaurus		Schnittmengen IZ-Klas- sifikation/ IZ-Thesaurus ¹⁰	
Notation	Text		Notation	Text	Notation	Text	Verhältnis Schnittmen- ge zu Klass.	Verhältnis Schnittmenge zu Thes.
DS	Sozialpädagogik, Sozialarbeit	= (hoch)	20600	Sozialarbeit und Sozialpädagogik	3.1...	Sozialwesen und Sozialpraxis	72%	12%
MS 6530	Soziologie der Freizeit	< (hoch)	20400	Freizeitforschung und Freizeitsoziologie	3.6...	Freizeitforschung, Freizeitsoziologie	89%	33%
MR 2000- MR 2950	Methoden der Sozialforschung	> (hoch)	10104	Forschungsarten der Sozialforschung	0.1.01	Forschungsarten	20%	7%
ME	Politische Systeme: Allgemeines u. Vergleich	>< (hoch)	10503	Staat, politisches System	4.4.01	Herrschaftsformen, Herrschaftssysteme, Herrschaftsstrukturen	51%	6%
DF	Systematische Pädagogik	< (mittel)	10601	Allgemeines, spezielle Theorien und „Schulen“, Methoden und Geschichte der Erziehungswissenschaft	0.2.02	Pädagogik (Theorien und Ansätze)	60%	4%
AN	Buch- und Bibliotheks- wesen, Informationswissenschaft	> (mittel)	1080500	Informationswissenschaft	0.3.01	Kommunikations- wissenschaft	20%	0,3%
MB 3250	Politik und Soziologie	>< (mittel)	10504	Politische Willensbildung, politische Soziologie	4.4.06	Politische Meinungs- und Willensbildung	26%	46%
PG	Rechtspflege, Gerichtsverfassung, ZivilVerfahrensrecht und Nebengebiete	< (niedrig)	40102	Justiz	4.5.02	Rechtswesen, Justizverwaltung	34%	6%
XF 1218	Medizinische Ausbildung, medizinische Psychologie, medizinische Soziologie	> (niedrig)	10708	Psychische Störungen, Behandlung und Prävention	3.4.02	Prävention, Therapie, medizinische Rehabilitation	59%	14%
AL 41500	Hochschul- und Universitätswesen: Wirtschafts-, Sozial- und politische Wissenschaften	>< (niedrig)	10202	Lehre und Studium, Professionalisierung und Ethik, Organisationen und Verbände der Soziologie	0.3.07	Soziologie/ Sozialpsychologie	84%	1,2%

¹⁰ Die erste Spalte enthält die Schnittmenge der Treffer der IZ-Klassifikation mit den Treffern der IZ-Thesaurus-Notation, ausgedrückt als prozentualer Anteil der Klassifikationstreffer insgesamt. Die zweite Spalte setzt diese Schnittmenge in Beziehung zu der Gesamtzahl der Treffer, die mit der Thesaurus-Notation erzielt wurden (ebenfalls ausgedrückt als prozentualer Anteil der Schnittmenge). Zur Illustration sei dies an einem Beispiel dargestellt (siehe Zeile 1): die Suche mit der IZ-Klassifikation 20600 ergab 5094 Treffer, die Suche mit der Thesaurus-Notation 3.1 ergab 31444 Treffer. Die Schnittmenge betrug 3687 Treffer. 3687 sind 72% von 5094 und 12% von 31444.

Übersicht: Testverknüpfungen von der Basis-Klassifikation über die IZ-Klassifikation zur IZ-Thesaurus-Notation

Basis-Klassifikation			IZ Klassifikation		IZ-Thesaurus		Schnittmengen IZ-Klassifikation/ IZ-Thesaurus ¹¹	
Notation	Text	Beziehung (Relevanz)	Notation	Text	Notation	Text	Verhältnis Schnittmenge zu Klass.	Verhältnis Schnittmenge zu Thes.
71.35	Kindersozio- logie, Jugendsoziologie	= (hoch)	10210	Jugendsoziologie	2.6.00	Lebensalter	86%	26%
71.00	Soziologie: Allgemeines	< (hoch)	10201	Allg. Soziologie, Makro- soziologie, spez. Theorien u. „Schulen“, Entw. u. Gesch.der Soziologie	0.3.07	Soziologie/ Sozialpsychologie	23%	17%
70.03	Methoden, Techniken und Organisation der sozialwissenschaft- lichen Forschung	> (hoch)	10104	Forschungsarten der Sozialforschung	0.1.01	Forschungsarten	20%	7%
71.11	Gesellschaft	>< (hoch)	10203	Soziologie von Gesamt- gesellschaften	1.2.00	Gesellschaftsordnun- gen, Gesellschaftssysteme	54%	6%
15.06	Politische Geschichte	< (mittel)	30301	Allgemeine Geschichte	0.3.09	Sonstige Disziplinen und Teilbereiche	13%	2%
85.05	Betriebssoziologie, Betriebspsychologie	> (mittel)	10205	Wirtschaftssoziologie	4.6...	Wirtschaft	88%	8%
05.20	Kommunikation und Gesellschaft	>< (mittel)	1080400	Massenkommunikation	4.2...	Kommunikation, Sprache, Publizistik	99%	3%
02.00	Wissenschaft und Kultur allgemein	< (niedrig)	10102	Wissenschaftstheorie, Wissenschaftsphilosophie, Wissenschaftslogik	0.1.00	Wissenschaftstheorie, Methodologie	63%	14%
71.80	Sozialpolitik: Allgemeines	> (niedrig)	11002	Lehre und Studium, Professionalisierung und Ethik, Organisationen und Verbände der Sozialpolitik	3.1.00	Sozialpolitik	50%	0,15%
83.66	Agrarwirtschaft	>< (niedrig)	10212	Agrarsoziologie	4.6.04	Wirtschaftszweige, -sektoren, -bereiche	66%	5%

¹¹ Die erste Spalte enthält die Schnittmenge der Treffer der in der gleichen Zeile genannten IZ-Klassifikations-Notation mit den Treffern der daneben genannten IZ-Thesaurus-Notation, ausgedrückt als prozentualer Anteil an den Treffern der Klassifikations-Notation insgesamt. Die zweite Spalte setzt die Schnittmenge in Beziehung zu den Treffern der Thesaurus-Notation (ebenfalls ausgedrückt als prozentualer Anteil der Schnittmenge). Zur Illustration sei dies an einem Beispiel dargestellt (siehe Zeile 1): Die Suche mit der IZ-Klassifikation 10210 ergab 8917 Treffer, die Suche mit der Thesaurus-Notation 2.6.00 ergab 29890 Treffer. Die Schnittmenge betrug 7686 Treffer. 7686 sind 86% von 8917 und 26% von 29890.

1.7 Personal

(in dieser Version nicht enthalten)

1.8 Arbeitstreffen

(in dieser Version nicht enthalten)

1.9 Workshops und Präsentationen

- 13. - 14.01.2000: CARMEN Workshop, Universität Osnabrück
- 24.02.2000: Organisation des Workshops "Probleme der Integration und der Kreuzreferenzierung von unterschiedlichen bzw. mehrsprachigen Thesauri und Klassifikationen im Kontext Virtueller Bibliotheken", Bonn
- 14.03.2000: Nase, Annemarie: "CARMEN AP12". Workshop 'DFG-Projekte zu Virtuellen Fachbibliotheken, AG Sacherschließung', Hamburg
- 11. - 12.05.2000: Organisation des Workshops 'CARMEN AG Heterogenität', ein gemeinsamer Workshop der AP 7, 11 und 12, Bonn
- 11.05.2000: Nase, Annemarie: "Crosskonkordanzen von Thesauri". Workshop 'CARMEN AG Heterogenität', Bonn
- 10. - 11. Oktober 2000: Posterpräsentation, Global Info-Workshop, Frankfurt am Main
- 17.11.2000: Schott, Hannelore: Vorführung von SIS-TMS anhand der Crosskonkordanzen von Thesauri (AP12). GESIS-Kuratoriumssitzung, Bonn
- 28.11.2000: Schott, Hannelore: Vorführung von SIS-TMS anhand der Crosskonkordanzen von Thesauri (AP12). Workshop der Vlib-Arbeitsgruppe, Hamburg
- 05.04.2001: Woldering, Britta: Crosskonkordanzen zwischen kontrollierten Indexierungssprachen – Projekt CARMEN AP12“, Bibliothekartag in Bielefeld
- 10. - 12.10.2001: Stand zu CARMEN im IBLC-Forum, Buchmesse, Frankfurt am Main
- 12.10.2001: Woldering, Britta: Konkordanzen zwischen Klassifikationen und Thesauri: Neue Möglichkeiten der Suche im Projekt CARMEN AP 12, Unterarbeitspaket Thesauri, Buchmesse in Frankfurt am Main
- 12.10.2001: Schöder, Albert: Konkordanzen zwischen Klassifikationen und Thesauri: Neue Möglichkeiten der Suche im Projekt CARMEN AP 12, Unterarbeitspaket Klassifikatione, Buchmesse in Frankfurt am Main
- 01.02.2001: Schott, Hannelore: Crosskonkordanzen von Thesauri: Ergebnisse und Perspektiven, CARMEN middleOfTheRoad Workshop, Informationszentrum Sozialwissenschaften, Bonn
- 18.01.2002: Woldering, Britta und Hannelore Schott: Heterogenitätsbehandlung bei Thesauri, CARMEN nextStep Workshop, Universität Osnabrück
- 18.01.2002: Schröder, Albert: Heterogenitätsbehandlung bei Klassifikationen, CARMEN nextStep Workshop, Universität Osnabrück
- 10.04.2002: Schröder, Albert: Konkordanzen zwischen Klassifikationen - neue Möglichkeiten der Suche im Projekt CARMEN, Bibliothekartag in Augsburg

1.10 Veröffentlichungen

- 17.08.2000, Geißelmann, Friedrich: The indexing of electronic publications - ways out of heterogeneity. International Federation of Library Associations and Institutions. 66th General Conference Jerusalem Workshop <<http://www.ifla.org/IV/ifla66/papers/173-181e.htm>>.
- 02.09.2000, Geißelmann, Friedrich: Die Erschließung elektronischer Publikationen. Wege aus der Heterogenität. Österreichischer Bibliothekartag Wien <<http://bt2000.univie.ac.at/abstracts/Geisselmann/Geisselmann.htm>>.
- 67. IFLA General Conference in Boston. Veranstaltungen der Division IV Bibliographic Control/Friedrich Geißelmann, Renate Gömpel, Christel Hengel, Martin Kunz, Monika Münnich, Werner Stephan, Claudia Werner. In: Bibliotheksdienst 35. Jg. (2001), H. 11 (1492-1511, hier S. 1507-1509)

Laufend aktuelle Informationen zum Projekt auf der Homepage zu CARMEN AP 12 unter <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html>