

Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung in Informationssystemen

*Anne-Kathrin Walter, Philipp Mayr, Maximilian Stempfhuber,
Arne Ballay*

Informationszentrum Sozialwissenschaften, Bonn



Dieses Dokument wird unter folgender creative commons Lizenz veröffentlicht: <http://creativecommons.org/licenses/by-nc-nd/2.0/>

Zusammenfassung

Die nutzerfreundliche Zusammenführung von Informationsangeboten erfordert nicht nur die Kopplung der Einzelangebote auf technischer Ebene – erst die semantische Integration der in der Regel unterschiedlich strukturierten und erschlossenen Information ermöglicht auf der Nutzerseite einen einheitlichen und effizienten Zugriff. Als Ergänzung zur Standardisierung von Metadaten und Inhaltserschließungsverfahren, die sich vor allem beim Neuaufbau von Informationssammlungen anbietet, spielen Crosskonkordanzen eine zentrale Rolle bei der Vernetzung von kontrollierten Erschließungsvokabularen unterschiedlicher Datenbanken oder Bibliothekskataloge. Erst deren Transferleistung – die Abbildung von Termen eines Thesaurus auf die Terme eines anderen Thesaurus durch semantische Relationen – ermöglicht es Nutzern, mit Begriffen eines Vokabulars gleichzeitig in mehreren Datenbanken zu recherchieren, ohne die Suchanfrage mehrfach und datenbankspezifisch zu formulieren. Am Beispiel des Informationsverbunds infoconnex¹ wird das Konzept der Crosskonkordanzen erläutert und anhand von konkreten Beispielen gezeigt, welche Möglichkeiten sie bei der integrierten Recherche in Fachdatenbanken aus Pädagogik, Sozialwissenschaften und Psychologie bieten.

1 Einleitung

Ein Blick auf die aktuelle Informationslandschaft zeigt eine Diversifizierung in viele dezentrale Informationsangebote, die online und zumeist unverbunden zur

¹ Das Projekt infoconnex wurde gefördert vom Bundesministerium für Bildung und Forschung (BMBF) und von der Deutschen Forschungsgemeinschaft (DFG).

Verfügung stehen. Mit zunehmender Vernetzung und interdisziplinärer Ausrichtung von Forschungsaktivitäten sehen sich Nutzer immer häufiger in der Situation, auf mehrere Informationsangebote zugreifen zu müssen, um ihr Informationsbedürfnis zu befriedigen. Die alleinige Nutzung z. B. des Online-Katalogs (OPAC) der lokalen Bibliothek führt genauso wenig zu einem befriedigenden Ergebnis wie die Beschränkung auf eine einzelne Fachdatenbank. Mit jedem Wechsel zwischen Informationsangeboten sieht sich der Nutzer aber vor das Problem gestellt, die eben bereits für den OPAC formulierte Suche für eine oder mehrere Fachdatenbanken neu zu formulieren.

Nicht nur die unterschiedlichen Benutzungsoberflächen bedingen eine erneute Einarbeitung in jedes der Recherchesysteme, auch die Schlagwortsuche, die Suche über das zur inhaltlichen Beschreibung von z. B. Publikationen verwendete, kontrollierte Vokabular, erfordert die wiederholte Auswahl von Suchbegriffen, die das Informationsbedürfnis des Nutzers möglichst gut beschreiben. Aufgrund der unterschiedlichen Größe, Spezifität und Fachbezogenheit der Vokabulare (z. B. Schlagwortnormdatei der Deutschen Bibliothek als großes, allgemeines Vokabular im Gegensatz zu einem relativ kleinen, aber fachspezifischen Thesaurus) können die für eine bestimmte Datenbank gewählten Suchbegriffe nur in seltenen Fällen direkt zur Suche in einer anderen Datenbank verwendet werden. Neben für den Nutzer relativ leicht zu behebenden Unterschieden in der Ansetzung der Thesaurusbegriffe (z. B. Ansetzung „Kind“ im Singular vs. Ansetzung „Kinder“ im Plural) erfordern vor allem die Verwendung unterschiedlicher Begriffe für das gleiche semantische Konzept (z. B. „Kommunikationsgeschichte“ in der Pädagogik vs. „Mediengeschichte“ in den Sozialwissenschaften) oder die Kombination mehrerer Begriffe (z. B. „Medien + Geschichte“ in der Psychologie) eine intensive Einarbeitung in den Sprachgebrauch der Fachgebiete oder Datenbanken. Nur wenige Nutzer dürften bereit und in der Lage sein, diesen aufwändigen Transfer bei der täglichen Informationsbeschaffung zu leisten – zumindest legen neuere Untersuchungen dies nahe [s. Boekhorst et al. 2003, Poll 2004]. Was aber ist zu tun, um den Nutzern den bequemen Zugriff auf die benötigten Informationen zu ermöglichen?

Eine Strategie, die bereits erste positive Resultate zeigt, ist die Zusammenführung bislang unverbundener Angebote. So bündeln die von der Deutschen Forschungsgemeinschaft (DFG) geförderten Virtuellen Fachbibliotheken und die vom Bundesministerium für Bildung und Forschung finanzierten Informationsverbünde bereits relevante Informationen der betreffenden Fächer. Darüber hinaus führt das fachübergreifende Wissenschaftsportal *vascoda*² beide Förderlinien zusammen und schafft einen zentralen Zugang zu wissenschaftlicher Information über Fächergrenzen hinweg. Für den Nutzer ist dadurch mehr – bislang verstreute – Information an einer Stelle zugänglich und der Aufwand beim

2 <http://www.vascoda.de>

Wechsel zwischen Informationsangeboten sinkt deutlich. Auf der qualitativen Ebene wurde das Potential der Integration aber bislang noch wenig genutzt. Zwar bieten viele Portale eine integrierte Suche (Metasuche) an, häufig werden die Suchbegriffe des Nutzers aber ohne weitere Behandlung an die angeschlossenen Datenbanken und OPCAs weitergeleitet – mit allen negativen Konsequenzen für die Größe (Recall) und Qualität (Precision) des Ergebnisses.

Die qualitative Verbesserung der integrierten Suche in Datenbanken unterschiedlicher Fächer hat sich in diesem Umfeld erstmalig der Informationsverbund Pädagogik – Sozialwissenschaften – Psychologie (infoconnex³) zum Ziel gesetzt. Dabei wurde versucht, der Heterogenität zu begegnen, die sich aus der Zusammenführung verschiedener Informationsangebote auf mehreren Ebenen ergibt (s. Stempfhuber 2002 und Strötgen 2004):

- Technische Heterogenität aufgrund der Zusammenführung unterschiedlicher Datenbanksysteme.
- Strukturelle Heterogenität aufgrund unterschiedlicher Modellierung der in den Datenbanken beschriebenen Entitäten (z. B. Literatur, Internetquellen, Institutionen, Lehrmaterialien) durch jeweils eigene Attributmengen (Metadatenschemata).
- Semantische Heterogenität aufgrund unterschiedlicher Inhaltserschließungssysteme, durch die die zulässigen Attributausprägungen definiert werden (z. B. Thesauri, Klassifikationen, Dokumentarten, Personenrollen, Länderlisten und andere Normdaten).

Der Fokus von infoconnex lag auf der Behandlung von semantischer Heterogenität, die durch die Verwendung jeweils eigener Attributausprägungen für semantisch äquivalente Attribute entsteht. Im Kontext von Literaturinformation betrifft dies häufig die Attribute „Schlagwort“ und „Klassifikation“, die zur Beschreibung der wesentlichen Inhalte von Publikationen verwendet werden und zwischen OPACs und Fachdatenbanken sehr starke Unterschiede aufweisen. So ist jede der Fachdatenbanken in infoconnex mit einem eigenen Thesaurus erschlossen, der sich hinsichtlich Struktur, Scope und Spezialisierungsgrad wesentlich von den anderen unterscheidet. Um dem Benutzer die Möglichkeit zu geben, mit einem beliebigen – in der Regel dem ihm am besten bekannten – Thesaurus eine Anfrage zu formulieren, die dann auf die drei unterschiedlich erschlossenen Fachdatenbanken angewendet wird, wurden die Thesauri bilateral durch so genannte Crosskonkordanzen miteinander verknüpft. Die Menge der zwischen den Begriffen unterschiedlicher Thesauri definierten Beziehungen (Relationen) erlaubt es im Rechercheprozess dann, die Anfrage des Nutzers vom Vokabular des Ausgangsthesaurus automatisch auf andere Vokabulare abzubilden und mit den für die Zieldatenbank geeigneten Begriffen zu suchen.

3 <http://www.infoconnex.de>

Im Folgenden werden zunächst mögliche Verfahren zur semantischen Heterogenitätsbehandlung vorgestellt und dann vertieft auf Crosskonkordanzen eingegangen. Am Beispiel der fachübergreifenden Recherche im Informationsverbund infoconnex wird dann die Anwendung der Crosskonkordanzen in der Praxis erläutert und ihre Wirksamkeit zur Vokabularerweiterung belegt.

2 Semantische Heterogenitätsbehandlung

Ein zentrales Problem des Information Retrieval ist die Vagheit der Abbildung zwischen der Anfrage eines Benutzers und den Indexierungsbegriffen, die den Inhalt – häufig nur die wahrscheinlich für Nutzer relevanten Aspekte – der im Informationssystem gespeicherten Dokumente beschreiben (siehe Abbildung 1: V1). Konkret bedeutet dies, dass ein Benutzer in seiner Anfrage für den gleichen semantischen Aspekt andere Begriffe verwendet, als – im Falle intellektueller Erschließung – die für die Inhaltserschließung verantwortliche Person vergeben hat. Auch bei der automatischen Indexierung tritt dieses Problem auf, allerdings hier zwischen dem Wortschatz des Informationssuchenden und den Autoren der Dokumente. Traditionell wird die Vagheit zwischen Anfrage- und Dokumentenebene durch Verfahren zur Termerweiterung behandelt, was „manuell“ durch den Nutzer geschehen kann (z. B. durch Verwendung eines Thesaurus mit Synonym- oder Ober-/Unterbegriffsbeziehungen) oder in ähnlicher Weise auch durch das Informationssystem (teil-)automatisch geschehen kann (z. B. durch Vorschläge für zusätzliche und alternative Suchbegriffe).

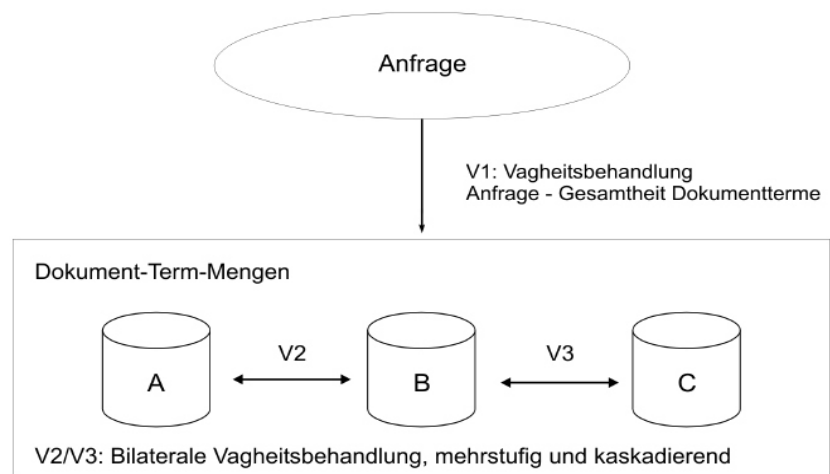


Abb. 1: Vagheitsbehandlung (aus Krause 2003)

Handelt es sich bei den zu durchsuchenden Datenbeständen um homogen erschlossene Datenbanken, sind diese Verfahren ausreichend. Anders bei heterogenen Dokumentenbeständen: Durch die unterschiedlichen Thesauri entsteht Vagheit bereits auf der inhaltlichen Beschreibungsebene der Dokumente. Stellt man die Datenbanken einander gegenüber und vergleicht, mit welchen Begriffen gleiche Dokumente oder solche mit ähnlichem Inhalt erschlossen sind, so stellt man fest, dass für gleiche semantische Konzepte je nach Erschließungsvokabular unterschiedliche Indexierungsbegriffe verwendet wurden, die sich in ihrer semantischen Präzision unterscheiden können. Die Konsequenz ist, dass die Anfrage des Nutzers nicht nur auf ein Indexierungsvokabular, sondern parallel auf mehrere abgebildet werden muss.

In der Praxis würde dies bedeuten, dass zwischen dem Vokabular, das der Nutzer zur Formulierung der Anfrage benutzt (z. B. den ihm gut bekannten Thesaurus einer Fachdatenbank) und jedem der Vokabulare der anderen Datenbanken eine Abbildungsvorschrift (z. B. eine Konkordanz) vorliegen muss. Dies ist häufig aber nicht der Fall. Als Ausweichstrategie kann die Abbildung der Nutzeranfrage bilateral und kaskadierend zwischen den beteiligten Vokabularen erfolgen, so dass die Anfrage zunächst auf ein (zentrales) Vokabular abgebildet und von dort aus schrittweise auf die anderen propagiert wird. Die Abbildung 1 verdeutlicht dies. Die Anfrage des Nutzers wird hier zunächst auf das Vokabular der Datenbank B abgebildet und erst von dort aus auf die Vokabulare der Datenbank A (Transformation V2) und Datenbank B (Transformation V3). Dabei ist konzeptionell auch eine Kombination von unterschiedlichen Verfahren bei den einzelnen Schritten zulässig.

Zur Behandlung der Vagheit bzw. der Heterogenität der Dokumentenbestände gibt es verschiedene Verfahren. Die zwei wesentlichen werden im Folgenden kurz vorgestellt.

- Intellektuell erstellte Crosskonkordanzen zwischen Klassifikationen und Thesauri

Crosskonkordanzen sind gerichtete, relevanzbewertete Relationen zwischen Termen zweier Thesauri oder Klassifikationen. Sie ermöglichen eine „Übersetzung“ von Anfragetermen eines Erschließungssystems in ein anderes. Die Erstellung der Relationen erfolgt intellektuell, d.h. ein Terminologie-Experte vergleicht verschiedene Begriffssysteme und setzt die Terme zueinander in Beziehung. Die Verknüpfung erfolgt dabei nach einem Verfahren, das im Projekt CARMEN AP12⁴ definiert wurde und folgende Arten von Relationen zwischen Termen unterschiedlicher Vokabulare festlegt:

- Eine Äquivalenzrelation wird für identische, synonyme oder quasi-synonyme Terme verwendet.

4 siehe <http://www.bibliothek.uni-regensburg.de/projects/carmen12/index.html.de>

- Eine Oberbegriffsrelation wird von einem semantisch engeren zu einem semantisch weiteren Term gesetzt.
- Eine Unterbegriffsrelation wird von einem semantisch weiteren zu einem semantisch engeren Begriff gesetzt.
- Die Ähnlichkeitsrelation wird für verwandte Begriffe gesetzt.

Das Verfahren ist mit einem nicht zu vernachlässigenden intellektuellen Aufwand verbunden und betrachtet die Terme der Vokabulare unabhängig von ihrem konkreten Gebrauch bei der Indexierung von Dokumenten.

- Quantitativ-statistische Ansätze

Bei diesen Verfahren werden die semantischen Relationen mit Hilfe von statistischen Methoden automatisch erzeugt. Grundlage hierfür sind Parallelkorpora, d.h. Bestände, deren Dokumente gleichzeitig nach zwei Sacherschließungssystemen indexiert sind. In einer „Kookkurrenzanalyse“ werden im Parallelkorpus die gemeinsam vorkommenden Terme aus jedem Vokabular auf Dokumentebene ausgewertet und daraus Wahrscheinlichkeiten für ein gemeinsames Vorkommen abgeleitet. Es werden also solche Gruppen von Termen beider Thesauri ermittelt, die besonders häufig bei der Indexierung von Dokumenten zusammen auftreten.

Die Schwierigkeit bei diesem Verfahren ist, einen geeigneten Parallelkorpus bereitzustellen. In Strötgen 2004, Binder et al. 2002 und Hellweg et al. 2001 werden Möglichkeiten zur Simulation von Parallelkorpora beschrieben. Anders als bei den Crosskonkordanzen kann bei statistischen Verfahren nicht nach unterschiedlichen Relationstypen differenziert werden, die erzeugten Relationen spiegeln jedoch den Gebrauch beider Vokabulare bei der Indexierung konkreter Dokumente wider.

3 Einsatz von Crosskonkordanzen am Beispiel von infoconnex

3.1 Crosskonkordanzen im Informationsverbund infoconnex

infoconnex ist ein interdisziplinärer Informationsverbund für die Fächer Pädagogik, Sozialwissenschaften und Psychologie, der die integrierte Recherche in den zentralen Fachdatenbanken der drei Disziplinen ermöglicht und auch in das deutsche Wissenschaftsportal *vascoda*⁵ integriert ist. Im Rahmen von infoconnex kamen erstmalig die in Kapitel 2 erläuterten Crosskonkordanzen im Echtbetrieb zum Einsatz.

⁵ siehe <http://www.vascoda.de/>

Das Projekt infoconnex wurde zwischen Juni 2001 und Juli 2005 durch das BMBF und die DFG gefördert und als Gemeinschaftsprojekt vom Informationszentrum Sozialwissenschaften (IZ), dem Deutschen Institut für Internationale Pädagogische Forschung (DIPF) und dem Zentrum für Psychologische Information und Dokumentation (ZPID) sowie den entsprechenden Sondersammelgebietsbibliotheken (Universitätsbibliothek Erlangen-Nürnberg, SSG Bildungsforschung; Stadt- und Universitätsbibliothek Köln, SSG Sozialwissenschaften; Saarländische Universitäts- und Landesbibliothek, SSG Psychologie) ins Leben gerufen. Ziel war es, ein qualitativ hochwertiges, durch Verfahren der Heterogenitätsbehandlung gestütztes und interdisziplinär angelegtes Fachinformationsportal zu entwickeln, dessen Literaturnachweisdatenbanken auch mit Volltexten kommerzieller Verlage verknüpft werden sollten:

- Fachinformationssystem (FIS) Bildung (Pädagogik),
- PSYINDEX (Psychologie),
- SOLIS (Sozialwissenschaften).

Der Gesamtbestand umfasst ca. eine Million Nachweise. Neben der fachspezifischen Suche in den einzelnen Fachdatenbanken ermöglicht infoconnex zusätzlich eine fachübergreifende Suche (siehe Abb. 2).

The screenshot shows the 'infoconnex' search portal. At the top, there are three tabs: 'Pädagogik' (green), 'Sozialwissenschaften' (red), and 'Psychologie' (blue). Below the tabs is a search form with the query 'Schlagwort='ökonomische theorie'-> (ökonomie und theorien)[28]'. The results are displayed in a table with columns for 'Jahr', 'Titel', 'Datenbank', and 'Verfügbarkeit'. The results are filtered by 'Pädagogik (121)', 'SOLIS (1657)', and 'Psychologie (28)'. The table shows five results, all from the 'PSYINDEX' database.

	Jahr	Titel	Datenbank	Verfügbarkeit
<input type="checkbox"/>	1983	Marketing ohne verhaltenswissenschaftliche Fundierung? [...]	PSYINDEX	6 S P X
<input type="checkbox"/>	1979	Theorie und Erfahrung : Beitrage zur Grundlagenproblematik der Sozialwissenschaften [...]	PSYINDEX	6 S P X
<input type="checkbox"/>	1983	Wirtschaftskriminalitaet als Prozess kollektiver Selbstschaedigung? : Ein Beispiel und einige theoretische Ueberlegungen [...]	PSYINDEX	6 S P X
<input type="checkbox"/>	1982	Ergebnisse der experimentellen Wirtschaftsforschung zum Optimierungsverhalten [...]	PSYINDEX	6 S P X
<input type="checkbox"/>	1978	Nicht-intendierte soziale Folgen individueller Handlungen [...]	PSYINDEX	6 S P X

Abb. 2: Fachübergreifende Schlagwortsuche mit Einsatz von Crosskonkordanzen in infoconnex

Da jede der drei Datenbanken mit einem eigenen Thesaurus erschlossen ist, hätte eine Schlagwortsuche ohne den Einsatz von Verfahren zur Heterogenitätsbehandlung zur Folge, dass nur mit Suchbegriffen, die in allen drei Thesauri vor-

handen sind, auch in allen drei Datenbanken Treffer gefunden würden. Alle semantischen Konzepte, die in den drei Thesauri zwar vorkommen, dort aber sprachlich unterschiedlich ausgedrückt sind, könnten bei der Suche ohne zusätzliche intelligente Verfahren nicht entsprechend nachgewiesen werden.

Um dem zu begegnen, werden in infoconnex Crosskonkordanzen eingesetzt. Sie ermöglichen die Transformation der vom Nutzer eingegebenen Schlagwörter, die in der Regel aus dem ihm am besten bekannten Thesaurus stammen werden, in entsprechende Schlagwörter der anderen Thesauri. Damit wird es möglich, an jede der drei Datenbanken spezifische Anfragen zu stellen, ohne dass der Nutzer diesen intellektuellen Transfer selbst leisten muss. Die Abbildung 3 veranschaulicht das prinzipielle Vorgehen. Die an alle drei Datenbanken gerichtete Anfrage des Nutzers (Suchoption „infoconnex“) wird in einem ersten Schritt auf die im System vorhandenen Thesauri abgebildet. Es wird also überprüft, in welchen Thesauri die eingegebenen Schlagwörter enthalten sind. In einem zweiten Schritt wird ermittelt, zu welchen anderen Thesauri Crosskonkordanzrelationen existieren. Diese werden dann dazu benutzt, um für die Suche in der entsprechenden Datenbank die Originalbegriffe aus der Anfrage durch die über die Crosskonkordanzen transformierten Begriffe zu ersetzen und eine datenbankspezifische Anfrage zu erzeugen. Die Crosskonkordanzen liegen in infoconnex grundsätzlich bidirektional vor, d. h. die Thesauri sind immer paarweise und in beide Richtungen miteinander verbunden, so dass Transfers in jede Richtung stattfinden können. Bei einer größeren Zahl von Thesauri ist diese Art der Verknüpfung aufgrund der Vielzahl an Kombinationen möglicherweise nicht mehr zu leisten. Daher wird in Kapitel 4.2 ein alternatives Verfahren mittels indirekter Transfers dargestellt.

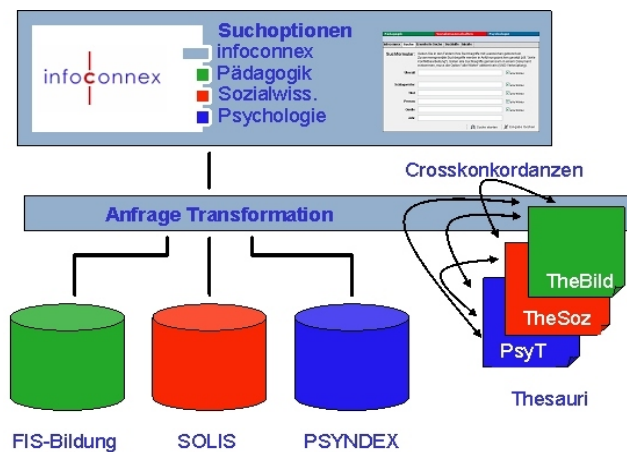


Abb. 3: Schematische Darstellung der Anfragetransformation in infoconnex

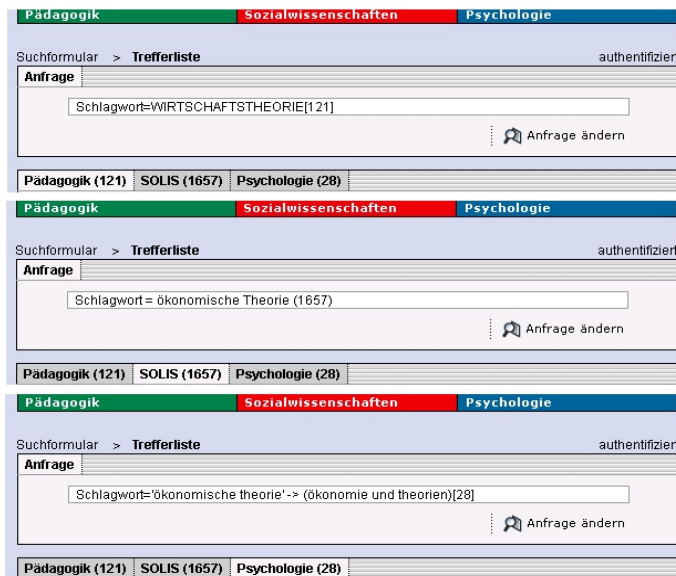


Abb. 4: Beispiel einer Anfragetransformation zum Schlagwort „ökonomische Theorie“

Die Abbildung 4 zeigt ein Beispiel dieser Anfragetransformation durch Crosskonkordanzen. Übergreifend über alle drei Datenbanken wird eine Schlagwortsuche nach „ökonomische Theorie“ durchgeführt. Im Thesaurus Sozialwissenschaften ist dieser Begriff ein gültiger Deskriptor. Die Anfrage „Schlagwort = ökonomische Theorie“ kann direkt an die Datenbank SOLIS weitergeleitet werden, wobei 1657 Dokumente gefunden werden. Im Thesaurus Psychologie hingegen ist dieser Begriff kein gültiger Deskriptor. Allerdings verweisen Crosskonkordanzrelationen auf äquivalente Terme in diesem Thesaurus, die stattdessen verwendet werden können. Die Anfrage des Nutzers wird für die Datenbank PSYINDEX zur Schlagwortkombination „Ökonomie + Theorien“ transformiert und es werden 28 Treffer gefunden. Auch im Thesaurus Pädagogik findet sich der Deskriptor „ökonomische Theorie“ nicht, eine Transformation zu „Schlagwort = Wirtschaftstheorie“ ergibt jedoch 121 Treffer in FIS Bildung.

Derzeit werden in infoconnex lediglich Äquivalenzrelationen genutzt, da sie in jedem Fall automatisch – also ohne Nutzerinteraktion – verwendet werden können. Die Verwendung weiterer Crosskonkordanz-Relationen (Oberbegriffe, Unterbegriffe, Ähnlichkeiten) soll dem Nutzer später über entsprechende Elemente auf der Benutzungsoberfläche schrittweise ermöglicht werden.

3.2 Weitere Beispiele für die Effekte von Crosskonkordanzen

Im Folgenden werden nun erste empirische Ergebnisse präsentiert, die zeigen, welchen Einfluss die Ausweitung des Suchvokabulars durch Crosskonkordanzen bei der datenbankübergreifenden Suche hat.

Beim Einsatz von Crosskonkordanzen stellt sich zunächst die Frage nach der Wirksamkeit dieser Rechercheunterstützung auf das Ergebnis. Wie sinnvoll der Einsatz von Crosskonkordanzen für die Literaturrecherche ist und wie Crosskonkordanzen aufgebaut werden müssen, um einen maximalen Effekt zu erzielen, kann letztendlich nur durch eine qualitative Bewertung der Ergebnisse durch Nutzer beantwortet werden, auf deren Basis dann detaillierte Messungen des Recall und der Precision erfolgen können. Diese Untersuchungen sind im Projekt „Kompetenzzentrum Modellbildung und Heterogenitätsbehandlung“⁶ geplant, das vom BMBF im Rahmen der „Kompetenznetzwerke Neue Dienste, Standardisierung, Metadaten“ gefördert wird.

Als vorbereitender Schritt für diese Untersuchungen wurde die Erhöhung der Trefferzahlen untersucht, die durch Äquivalenzrelationen (siehe Kap. 2) zwischen den drei kontrollierten Vokabularen von infoconnex erzielt werden können. Die in den Tabellen 1 bis 3 dargestellten Ergebnisse sind repräsentativ für Termtransformationen, für die in mindestens zwei der infoconnex-Datenbanken Treffer generiert werden können. Dabei werden nur solche Termtransformationen betrachtet, bei denen die Zielterme nicht zeichengleich mit dem Ausgangsterm sind, also eine „intelligente“ Transferleistung erbracht wurde (kursiv gekennzeichnet). In Tabelle 1 (Zeile 3) finden sich zwei dieser Termtransformationen. Der Deskriptor „berufliche Weiterbildung“ aus dem sozialwissenschaftlichen Thesaurus wird in Berufsbildung (Thesaurus Bildung) und „Weiterbildung im Beruf“ (Thesaurus Psychologie) transformiert.

Die Relevanz der zusätzlichen Treffer in den folgenden Darstellungen wurde nicht bewertet. Auf die Überlappung der drei kontrollierten Vokabulare und die Zusammensetzung der einzelnen in infoconnex eingesetzten Crosskonkordanzen wurde bereits an anderer Stelle hingewiesen (Mayr, Stempfhuber, Walter 2005).

Die folgenden Beispiele zeigen – ausgehend von Deskriptoren des Thesaurus Sozialwissenschaften TheSoz – die Relationen zu den Deskriptoren bzw. Deskriptorkombinationen in den beiden Vokabularen Thesaurus Bildung TheBild (CK) und Psyndex Terms PsyT (CK) sowie die Anzahl der Treffer dieser Konkordanzen in den drei Datenbanken SOLIS, FIS Bildung und Psyndex. Die Deskriptoren und Treffer, die ohne intellektuelle Konkordanzerstellung für die integrierte Recherche der drei Datenbanken nicht zur Verfügung gestanden hätten, sind in den folgenden Tabellen kursiv gekennzeichnet. Zeichengleiche Deskrip-

6 <http://www.gesis.org/Forschung/Informationstechnologie/komohe.htm>

toren spielen eine Sonderrolle (siehe unten) und werden nicht gesondert hervorgehoben.

Tabelle 1: Auswahl häufig zur Inhaltserschließung verwendeter TheSoz-Deskriptoren mit Termtransformationen in die Vokabulare TheBild und PsyT. Zusätzlich ist die Trefferanzahl angegeben, die mit dem Deskriptor bzw. den Transformationen in der jeweiligen Datenbank erreicht werden kann (Stand Anfang 2005).

TheSoz	SOLIS	TheBild (CK)	FIS Bildung	PsyT (CK)	Psyindex
Arbeitsloser	1.995	Arbeitsloser	1.432	Arbeitslosigkeit + Arbeiter	12
berufliche Weiterbildung	4.869	Berufsbildung	3.494	Weiterbildung im Beruf	529
Entwicklungsland	18.278	Entwicklungsland	2.756	Entwicklungsländer	111
Erfahrung	1.875	Erfahrung	1.727	Erfahrungen (Erlebnisse)	463
Gewerkschaftspolitik	2.193	Gewerkschaftspolitik	50	Gewerkschaften + Politik	2
Informationstechnologie	1.778	Informationstechnik	3.358	Information + Technologie	108
Kulturpolitik	1.599	Kulturpolitik	1.441	Kultur (Anthropologie) + Politik	12
Modellentwicklung	832	Modellentwicklung	231	Modelle + Entwicklung	154
ökonomische Theorie	1.613	Wirtschaftstheorie	91	Ökonomie + Theorien	28
politische Entwicklung	1.867	politische Entwicklung	18	Politik + Entwicklung	122
politische Kommunikation	1.383	Politische Kommunikation	15	Politik + Kommunikation	42
Prävention	1.624	Praeventivmaßnahme	2.526	Prävention	3.355
soziale Umwelt	1.044	Soziales Milieu	189	Soziale Umwelt	548
Technologiepolitik	1224	Technologiepolitik	113	Technologie + Politik	11

Tabelle 1 zeigt Deskriptoren des TheSoz (linke Spalte), die vergleichsweise häufig zur Indexierung von Dokumenten in SOLIS verwendet werden (siehe 2. Spalte, Treffer in SOLIS). Genuin sozial- bzw. politikwissenschaftliche Begriffe wie politische Entwicklung, politische Kommunikation, ökonomische Theorie oder Gewerkschaftspolitik liefern insbesondere in SOLIS deutlich mehr als 1.000 Dokumente. Dass auch in den beiden Datenbanken der anderen Fächer (FIS Bildung und Psyindex) durch die Crosskonkordanzen zusätzliche Doku-

mente gefunden werden, verdeutlicht ihren Mehrwert für interdisziplinäre Fragestellung. Auffällig ist, dass aufgrund der inhaltlichen Nähe des Thesaurus Sozialwissenschaften und des Thesaurus Bildung relativ viele Deskriptoren zeichengleich vorliegen⁷. Dies trifft z. B. auf folgende Deskriptoren zu: Arbeitsloser, Entwicklungsland, Erfahrung, Kulturpolitik usw. Die Überlappung zweier Thesauri durch zeichengleiche Terme, die sehr stark von der Größe und Spezialisierung der Vokabulare abhängt, kann beispielsweise als ein Indikator gewertet werden, ob es sinnvoll ist, eine Crosskonkordanz für diese zwei Vokabulare zu erstellen bzw. wie hoch der intellektuell zu leistende Aufwand sein wird.

Weiterhin fällt auf, dass die Überführung der TheSoz-Deskriptoren in das Vokabular des PsyT (CK) terminologisch schwieriger ist und folglich viel häufiger Deskriptorkombinationen verwendet werden müssen, um die Semantik der TheSoz-Deskriptoren auszudrücken. Beispiele hierfür sind Arbeitslosigkeit + Arbeiter, Gewerkschaft + Politik oder Modell + Entwicklung. Die Analyse der Crosskonkordanzen TheSoz→PsyT und TheSoz→TheBild in infoconnex bestätigt diesen Eindruck sehr deutlich.

Tabelle 2 zeigt als Kontrast ausgewählte Deskriptoren des TheSoz, die vergleichsweise selten zur Indexierung von Dokumenten in SOLIS verwendet werden. Beispielsweise liefert die Schlagwortsuche nach den Deskriptoren Biologieunterricht, Ethikunterricht oder Pharmakologie in SOLIS kaum Treffer, während in FIS Bildung sehr viele und in Psyndex zumindest einige zusätzliche Treffer gefunden werden. Dies zeigt, dass die in infoconnex erstellten Crosskonkordanzen über das gesamte Spektrum des Ausgangsthesaurus hinweg zu einer Erweiterung des Dokumentenraums auf zwei verwandte Fächer (Pädagogik und Psychologie) führen.

Die beiden Deskriptoren Biologieunterricht und Ethikunterricht müssen im Fall des PsyT durch Kombination von zwei Deskriptoren (Ethik + Lehren bzw. Naturwissenschaftlicher Unterricht + Biologie) ausgedrückt werden. Das Beispiel des TheSoz-Deskriptors schulpsychologischer Dienst zeigt, dass die Kombination zweier Deskriptoren nicht immer positive Effekte für eine Konkordanz bringt. Die Kombination der beiden PsyT-Deskriptoren Schulpsychologie + Soziale Dienste bringt im Fall Psyndex keine Treffer, was beim Scope dieser Datenbank überrascht. Recherchiert man lediglich mit dem Psyndex-Deskriptor Schulpsychologie, liefert Psyndex 280 Treffer. Folglich werden durch die Kombination mit dem Deskriptor Soziale Dienste relevante Dokumente ausgeschlossen.

⁷ Dies hat mit der Entstehungsgeschichte des Thesaurus Bildung zu tun, der sich aus den Vokabularen mehrerer Einzelthesauri zusammensetzt und dadurch zwangsläufig auch andere Fachgebiete überlappt. Der Thesaurus befindet sich zur Zeit in Überarbeitung und wird in seiner neuen Version diesen Grad der Überlappung mit dem Thesaurus Sozialwissenschaften wohl nicht mehr aufweisen.

sen⁸. Dieser negative Effekt intellektuell erstellter Crosskonkordanzen, insbesondere bei Termkombinationen, kann letztlich nur ausgeschlossen werden, wenn Relationen neben der semantischen Prüfung zusätzlich über den Recall in der Zieldatenbank geprüft werden. Diese Prüfung ließe sich relativ leicht automatisieren und könnte in die laufende Crosskonkordanzerstellung eingliedert werden.

Tabelle 2: Auswahl selten verwendeter TheSoz-Deskriptoren mit Termtransformationen in die Vokabulare TheBild und PsyT, sowie die Trefferanzahl der Deskriptoren bzw. Deskriptor-Kombinationen in den Datenbanken SOLIS, FIS Bildung & Psyndex (Trefferzahlen sind vom Stand Anfang 2005).

TheSoz	SOLIS	TheBild (CK)	FIS Bildung	PsyT (CK)	Psyndex
Berufsgewerkschaft	2	Gewerkschaft	2175	Berufe + Gewerkschaften	2
Biologieunterricht	7	Biologieunterricht	4815	Naturwissenschaftlicher Unterricht + Biologie	10
Ethikunterricht	5	Ethikunterricht	818	Ethik + Lehren	9
Facharbeiterberuf	71	Facharbeiter	853	Facharbeiter + Berufe	2
frauenspezifische Verfolgung	16	Frau + Verfolgung	3	Verfolgung + Frauen	2
Islamismus	64	Islamischer Fundamentalismus	4	Islam + Fundamentalismus	3
moralisches Urteil	23	Moralisches Urteil	12	Moral + Beurteilung	65
nichteheliches Kind	104	Uneheliches Kind	4	Uneheliche Kinder	20
Pharmakologie	8	Pharmazie	80	Pharmakologie	52
Piaget, J.	147	Piaget, Jean	230	Piaget (Jean)	208
Psychometrie	36	Psychometrie	74	Psychometrie	438
schulpsychologischer Dienst	13	Schulpsychologische Beratung	306	Schulpsychologie + Soziale Dienste	0
Softwareergonomie	14	Softwareergonomie	10	Software + Ergonomie	194
Täter-Opfer-Beziehung	189	Taeter-Opfer-Beziehung	14	Täter + Kriminalitätsoffer	77
Umweltethik	49	Oekologische Ethik	2	Umwelt + Ethik	18

Dass auch das Problem der Eigennamen – und insbesondere der mangelnden Standardisierung dieser – durch Crosskonkordanzen positiv beeinflusst werden kann, wird am Beispiel des TheSoz-Deskriptors Piaget, J. deutlich. Durch die zwischengeschaltete Crosskonkordanz können – obwohl alle drei Thesauri den

⁸ Die Suche nach dem Schlagwort Schulpsychologie kombiniert mit der Freitextsuche Dienst bringt in Psyndex 43 relevante Treffer.

Namen unterschiedlich ansetzen – in allen drei Datenbanken Publikationen des Entwicklungspsychologen Jean Piaget gefunden werden.

Die weiteren, in Tabelle 3 aufgezeigten Beispiele sind dem Systematikbereich 3.4 Medizin und Gesundheitswesen des Thesaurus Sozialwissenschaften entnommen.

Tabelle 3: TheSoz-Deskriptoren und Termtransformationen des Systematikbereichs 3.4 „Medizin und Gesundheitswesen“ (Trefferzahlen: Stand August 2005)

TheSoz	SOLIS	TheBild (CK)	FIS Bildung	PsyT (CK)	Psyn- dex
Krankheit	6450	Krankheit	1506	Erkrankungen	2998
öffentliches Gesundheitswesen	239	Gesundheitswesen	628	Gesundheits- versorgungssystem	1835
Arzt-Patient- Beziehung	485	-	-	Ärzte + Patienten	234
Ethnomedizin	72	-	-	Volksmedizin	204
körperliche Entwicklung	75	Koerperliche Entwicklung	160	Physische Entwicklung	109
gesetzliche Krankenversicherung	491	Krankenversicherung	144	Krankenversicherung von Arbeitnehmern	15
Knochenkrankheit	-	Knochenkrankheit	2	Osteoporose	13
Impfung	7	Impfung	33	Immunisierung	12
Kinderlähmung	5	Kinderlaehmung	7	Poliomyelitis	7
geriatrische Rehabilitation	1	-	-	Geriatrische Patienten + Rehabilitation	123

Anhand einiger Deskriptoren (z. B. Ethnomedizin, Knochenkrankheit, Kinderlähmung und geriatrische Rehabilitation), die in den Sozialwissenschaften sicherlich ausgesprochene Spezialgebiete repräsentieren und für die z. T. im Thesaurus Bildung keine Entsprechungen gefunden werden konnten, wird sehr gut deutlich, warum die Crosskonkordanzen in Verbindung mit der interdisziplinären Suche eine entscheidende Verbesserung für den Nutzer bewirken. Für die oben genannten Deskriptoren, die in der Datenbank SOLIS kaum Treffer generieren, bringt gerade die Crosskonkordanz zur Psychologie eine große Erhöhung der Trefferzahlen, die einem sozialwissenschaftlichen Nutzer als intelligente Mehrwertleistung automatisch angeboten werden kann.

3.3 Zusammenfassung

- Für den Recherchierenden bedeutet der Einsatz von Crosskonkordanzen eine deutliche Vergrößerung des Suchraums, den er von jedem einzelnen der verwendeten Thesauri aus nun erreichen kann.
- Erste Untersuchungen in infoconnex zeigten, dass sich durch Crosskonkordanzen eine signifikante Erweiterung des Suchvokabulars ergibt (im Schnitt um ca. 50 % gegenüber dem Ausgangsvokabular).
- Besonders im interdisziplinären Bereich konnte belegt werden, dass Crosskonkordanzen einen informationellen Mehrwert bieten, da sie Nischen eines Fachgebiets mit potentiell zentraleren Bereichen eines anderen Fachgebiets verbinden können.
- Die vorliegenden, rein quantitativen Ergebnisse lassen zwar noch keine präzisen Aussagen zur Relevanz der durch Crosskonkordanzen gelieferten, zusätzlichen Treffer zu, die erreichte Vergrößerung der Treffermenge stellt aber eine notwendige Voraussetzung für die angestrebte Verbesserung des Ergebnisses dar.

4 Weitere Einsatzmöglichkeiten von Crosskonkordanzen

Aufgrund der positiven Erfahrungen mit dem Einsatz von Crosskonkordanzen im Rahmen von infoconnex soll in den künftigen Ausbaustufen des Portals das Potential dieser semantischen Technologie in größerem Umfang genutzt werden. Dies betrifft vor allem den Einsatz aller in den Crosskonkordanzen enthaltenen Relationen sowie die Implementierung von indirekten Transformationen in Fällen, in denen ein Transfer zwischen zwei Thesauri nicht auf direktem Weg möglich ist.

4.1 Einsatz der zusätzlichen Relationstypen

Die bislang in infoconnex eingesetzten Äquivalenzrelationen stellen nur einen kleinen Teil der insgesamt in den Crosskonkordanzen enthaltenen Beziehungen zwischen Deskriptoren zweier Thesauri dar. Sie unterscheiden sich von den anderen Relationstypen (Oberbegriff, Unterbegriff, Ähnlichkeit) im Grad der Übereinstimmung der semantischen Konzepte, die verbinden.

Während bei Äquivalenzrelationen davon ausgegangen werden kann, dass die Präzision der Termtransformation so hoch ist, dass sie ohne Rückfrage automatisch für den Nutzer durchgeführt werden kann, ist dies bei den anderen Relationen nicht zwangsläufig der Fall. Besonders bei Oberbegriffsrelation wie zwischen Hochadel und Adel verringert sich die Präzision und es ist zu erwar-

ten, dass deutlich mehr und vor allem weiter vom Informationsbedürfnis des Nutzers entfernte Treffer nachgewiesen werden. Das Gegenteil ist bei Unterbegriffsrelation wie zwischen Kampagne und Werbekampagne zu erwarten, bei der Ähnlichkeitsbeziehung zwischen Konfliktverhalten und Konfliktfähigkeit ist die Auswirkung auf das Ergebnis – und vor allem die Bewertung durch den Nutzer – weitgehend offen. Weitere Untersuchungen werden hier erst zeigen müssen, ob bestimmte Relationen ebenfalls automatisch, also ohne explizite Aktivierung durch den Nutzer, ausgeführt werden können, ohne das Ergebnis zu verschlechtern. Eine wesentliche Rolle werden in jedem Fall die Menge der zusätzlichen Treffer, der Einsatz dieser Relationen für sich alleine oder in Kombination mit anderen und die Fachgebiete spielen (mehrere Fachgebiete im Gegensatz zum Einsatz zwischen Thesauri eines Fachgebietes). Daneben entscheidet auch die Form der Bereitstellung dieser Möglichkeiten für den Nutzer über ihre Akzeptanz. In Stempfhuber 2003 werden Gestaltungsideen vorgestellt, wie Crosskonkordanzen unterschiedlichen Nutzergruppen in jeweils spezifischer Form und interaktiv zur Rechercheformulierung angeboten werden können.

4.2 Indirekte Termtransformation

Werden mehrere Datenbanken integriert durchsucht, wird die Anfrage im Idealfall durch eine Crosskonkordanz direkt in alle beteiligten Erschließungssysteme übersetzt (direkter Transfer). Allerdings ist es nicht immer möglich und sinnvoll, sämtliche Erschließungssysteme untereinander zu verknüpfen, da die Erstellung von Crosskonkordanzen – je nach Größe der Vokabulare – mit nicht zu vernachlässigendem Aufwand verbunden ist. Dies gilt nicht nur für fachlich weit entfernte Thesauri und Klassifikationen wie Physik und Psychologie, auch innerhalb eines Faches kann der Aufwand zu groß sein. Beispielsweise besteht in den Sozialwissenschaften der Wunsch, die Datenbank SOLIS des IZ gemeinsam mit der Datenbank SoLit des Deutschen Zentralinstituts für soziale Fragen (DZI) und dem Bibliothekskatalog des Sondersammelgebiets (SSG) Sozialwissenschaften an der Universitäts- und Staatsbibliothek (USB) Köln zu durchsuchen. Letzterer ist mit der Schlagwortnormdatei (SWD) der Deutschen Bibliothek und SoLit mit dem Thesaurus des DZI verschlagwortet. Der Aufwand, jedes der drei Vokabulare paarweise miteinander zu verbinden, ist bereits relativ groß; zieht man in Betracht, dass dies nur ein kleiner Teil der in den Sozialwissenschaften relevanten Datenbanken ist, so wird schnell deutlich, dass eine paarweise bidirektionale Verknüpfung von Vokabularen nicht skaliert, mit vernünftigem Aufwand also nicht zu leisten ist.

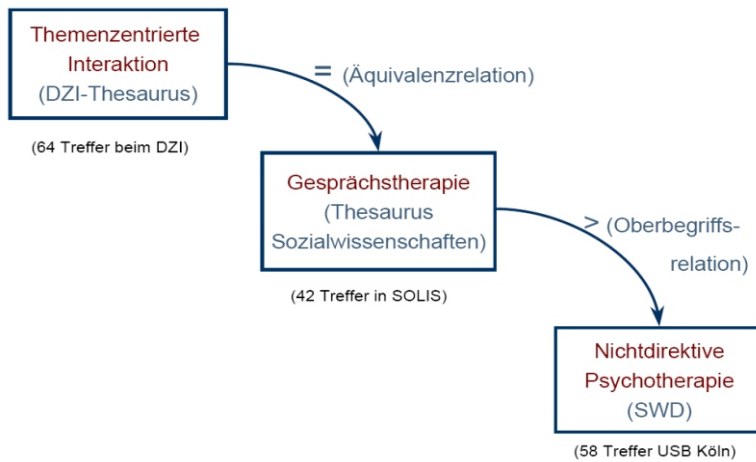


Abbildung 5: Indirekte Termtransformation (Beispiel)

Gelöst werden kann dieses Problem durch einen indirekten Termtransfer, also indem bei der Termtransformation ein „Umweg“ – oder Mehrschrittverfahren – gewählt wird. Dabei wird versucht, durch sequentielle Anwendung mehrerer Crosskonkordanzen einen Transfer zwischen zwei Erschließungssystemen zu erreichen, die nicht direkt miteinander verbunden sind. Abbildung 5 zeigt ein Beispiel, in dem ein Nutzer gleichzeitig in allen drei Datenbanken suchen möchte und dazu das Schlagwort themenzentrierte Interaktion aus dem DZI-Thesaurus verwendet. Im ersten Schritt – zur Suche in SOLIS – wird eine direkte Transformation (Gesprächstherapie) zum Thesaurus Sozialwissenschaften durchgeführt, woraus 42 Treffer resultieren. Für den zweiten Schritt – der Suche im Bibliothekskatalog – kann nun keine direkte Transformation vom DZI-Thesaurus aus durchgeführt werden, da eine derartige Crosskonkordanz nicht existiert. Stattdessen wird ein Zweischnittverfahren angewendet, in dem zunächst vom DZI-Thesaurus zum IZ-Thesaurus und erst von dort aus dann zur SWD transformiert wird (nichtdirektive Psychotherapie). Zur Sicherstellung einer möglichst hohen Präzision werden – wenn vorhanden – Äquivalenzrelationen verwendet. Da aber mit zunehmender Zahl an beteiligten Crosskonkordanzen dies nicht immer sichergestellt werden kann, dem Nutzer aber ein Ergebnis geboten werden soll, muss auch auf andere Relationen zurückgegriffen werden. Die theoretische Fundierung und praktische Erprobung von Mehrschrittverfahren stellt daher auch einen zukünftigen Forschungsschwerpunkt am IZ dar.

5 Ausblick

Crosskonkordanzen zwischen Erschließungsvokabularen, so haben erste Untersuchungen gezeigt, stellen ein probates Mittel dar, den Nutzer bei der gleichzeitigen Recherche in unterschiedlich erschlossenen Datenbanken zu unterstützen. Die Crosskonkordanzen leisten dabei automatisch den Transfer der Suchbegriffe aus der Anfrage des Nutzers in die unterschiedlichen Erschließungsvokabulare, mit denen die Datenbanken jeweils erschlossen sind. Der Nutzer wird in vielen Fällen vom entsprechenden kognitiven Aufwand befreit und kann sich intensiver der Bewertung der Rechercheergebnisse widmen. Entsprechende Benutzungsschnittstellen erlauben es dem fortgeschrittenen Nutzer, eine Vorschau auf das Ergebnis der Transformationen zu erhalten und diese sehr detailliert zu beeinflussen.

Der hier beschriebene Einsatz von Crosskonkordanzen bezog sich ausschließlich auf Textinformationen, also Literaturnachweisdatenbanken und Bibliothekskataloge. Doch auch im Bereich von Fakteninformationen, zum Beispiel bei statistischen Zeitreihen aus der Marktforschung, spielen sie eine Rolle, da hier ebenfalls eine Vielzahl von Datenbanken gemeinsam durchsucht werden müssen, um die benötigten Informationen zusammenzutragen (s. Stempfhuber et al. 2003). Da die statistischen Ämter weltweit in der Regel länderspezifische Nomenklaturen (also kontrollierte Vokabulare für Fakteninformation) zur Erschließung ihrer Statistiken verwenden, müssen gerade im internationalen Bereich häufig mehrere Nomenklatorsysteme verwendet werden, um die gewünschte Information zu erhalten.

Eine konzeptionelle Erweiterung erfährt der Einsatz von Crosskonkordanzen bei der gleichzeitigen Recherche von Text- und Faktendaten (s. Stempfhuber 2005). Hier gilt es nicht nur die Sprache der amtlichen Statistik auf die Sprache in Fachpublikationen abzubilden. Auch die grundsätzlich verschiedene Art der Inhaltserschließung von Texten und Fakten bereitet in manchen Fällen Probleme. Dies wird besonders dann deutlich, wenn nicht Aggregatdaten, sondern Primärdaten (z. B. sozialwissenschaftliche Umfragen) betrachtet werden. Die Komplexität der semantischen Konstrukte, die zur Inhaltserschließung von Primärdaten benötigt werden, um einem Nutzer nicht nur das Auffinden, sondern auch die korrekte Interpretation der Daten zu ermöglichen, geht weit über den Aufwand in der normalen dokumentarischen Praxis hinaus und lässt typische Thesauri oder Nomenklaturen nicht als geeignet erscheinen. Zukünftige Forschungsarbeiten werden daher verstärkt auch Ontologien mit einbeziehen, wenn es darum geht, Text- und Faktendaten zu dokumentieren und integriert recherchierbar zu machen (siehe Krause&Stempfhuber 2005).

Literatur

- Boekhorst, P.; Kayß, M.; Poll, R. (2003): Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung. Teil 1: Informationsverhalten und Informationsbedarf der Wissenschaft. Universitäts- und Landesbibliothek Münster und infas Institut für angewandte Sozialwissenschaft GmbH.
- Binder, G. et al.: Heterogenitätsbehandlung bei textueller Information verschiedener Datentypen und Inhaltserschließungsverfahren. IZ-Arbeitsbericht Nr. 24, 2002. S. 64 ff.,
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_24.pdf
- Hellweg, Heiko; Krause, Jürgen; Mandl, Thomas; Marx, Jutta; Müller, Matthias N.O.; Mutschke, Peter; Strötgen, Robert (2001): Treatment of Semantic Heterogeneity in Information Retrieval. Bonn: IZ Sozialwissenschaften. 47 S. (IZ-Arbeitsbericht; Nr. 23) URL:
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_23.pdf
- Krause, Jürgen (2003): Standardisierung von der Heterogenität her denken - Zum Entwicklungsstand bilateraler Transferkomponenten für digitale Fachbibliotheken. IZ-Arbeitsbericht Nr. 28, 2003. S.14 ff.,
http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_28.pdf
- Krause, Jürgen (2004): Konkretes zur These, die Standardisierung von der Heterogenität her zu denken. In: ZfBB: Zeitschrift für Bibliothekswesen und Bibliographie 51, Nr. 2, S. 76 - 89
- Krause, Jürgen; Stempfhuber, Maximilian (2005): Nutzerseitige Integration sozialwissenschaftlicher Text- und Dateninformationen aus verteilten Quellen, (01. Juli). - 6. Wissenschaftliche Tagung „Datenfusion und Datenintegration“, Statistisches Bundesamt Wiesbaden, 30. Juni - 01. Juli 2005
- Mayr, Philipp; Stempfhuber, Maximilian; Walter, Anne-Kathrin (2005): Auf dem Weg zum wissenschaftlichen Fachportal – Modellbildung und Integration heterogener Informationssammlungen. In: Schmidt, Ralph (Hrsg.): 27. DGI-Online-Tagung. Frankfurt am Main: DGI.
- Poll, Roswitha (2004): Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung Teil 1: Informationsverhalten und Informationsbedarf der Wissenschaft. In: ZfBB 51(2004) S. 59 – 75.
- Schott, Hannelore; Schroeder, Albert (2004): Crosskonkordanzen von Thesauri und Klassifikationen. S. 41 - 49. In: Budin, Gerhard; Ohly, H. Peter (Hrsg.):

- Wissensorganisation in kooperativen Lern- und Arbeitsumgebungen. Würzburg: Ergon. (Fortschritte der Wissensorganisation 8).
- Stempfhuber, Maximilian (2003): Objektorientierte Dynamische Benutzungsoberflächen – ODIN. Behandlung semantischer und struktureller Heterogenität in Informationssystemen mit den Mitteln der Softwareergonomie. Bonn: Informationszentrum Sozialwissenschaften. (Forschungsberichte; Band 6).
- Stempfhuber, Maximilian (2005): Text-Fakten-Integration in Informationssystemen. S. 263 - 274. Knowledge eXtended: die Kooperation von Wissenschaftlern, Bibliothekaren und IT-Spezialisten; 3. Konferenz der Zentralbibliothek; 2. - 4. November 2005; Vorträge und Poster. Jülich: Forschungszentrum, Zentralbibliothek. (Schriften des Forschungszentrums Jülich: Reihe Bibliothek/Library; 14).
- Stempfhuber, Maximilian; Hellweg, Heiko; Schaefer, André (2002): ELVIRA: User Friendly Retrieval of Heterogenous Data in Market Research. S. 299 - 304. In: Callaos, Nagib; Hernandez-Encinas, Luis; Yetim, Fahri (Hrsg.): SCI 2002: The 6th World Multiconference on Systemics, Cybernetics and Informatics; July 14 - 18, 2002, Orlando, USA; Proceedings, Vol. I: Information Systems Development I. Orlando: TPA Publ.
- Strötgen, Robert (2004): ASEMOS. Weiterentwicklung der Behandlung semantischer Heterogenität. In: Bekavac, Bernard; Herget, Josef; Rittberger, Mark (Hrsg.): Information zwischen Kultur und Marktwissenschaft; Proceedings des 9. Internationalen Symposiums für Informationswissenschaft (ISI 2004). Konstanz: UVK. (Schriften zur Informationswissenschaft; Bd. 42), S. 269 - 281.

Angaben zu den Autoren

Informationszentrum Sozialwissenschaften (IZ)
 Abt. Forschung und Entwicklung
 Lennéstraße 30
 53113 Bonn
 walter@bonn.iz-soz.de

Anne-Kathrin Walter studierte Informatik an der Universität Bremen. Seit September 2004 arbeitet sie im Projekt „Modellbildung und Heterogenitätsbehandlung“ am Informationszentrum Sozialwissenschaften in Bonn. Ihre Forschungsinteressen liegen im Information Retrieval, u.a. im Bereich der semantischen Heterogenitätsbehandlung.

Philipp Mayr studierte Bibliothekswissenschaft, Informatik und Soziologie an der Humboldt-Universität zu Berlin. Zu seinen Forschungsinteressen gehören

Information Retrieval im Bereich der Internet-Suchmaschinen sowie Metriken des Internet (Webometrie). Seit November 2004 arbeitet er im Projekt „Modellbildung und Heterogenitätsbehandlung“. Philipp Mayr ist wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften in Bonn.

Maximilian Stempfhuber studierte Informationswissenschaft und Wirtschaftsinformatik an der Universität Regensburg und promovierte in Informatik an der Universität Koblenz-Landau. Er ist stellvertretender Direktor des Informationszentrum Sozialwissenschaften in Bonn und leitet dort die Abteilung informationswissenschaftliche Forschung und Entwicklung. An der Universität Koblenz-Landau führt er im Rahmen von Lehraufträgen Veranstaltungen in der Informatik und Computervisualistik durch. Seine Forschungsschwerpunkte sind Softwareergonomie, Visualisierung und Information Retrieval.

Arne Ballay studierte Geografie, Schwerpunkt Geo-Informatik and der Rheinischen Friedrich-Wilhelms Universität Bonn. Seit November 2001 arbeitet er als wissenschaftlicher Mitarbeiter am Informationszentrum Sozialwissenschaften unter anderem im Projekt „Informationsverbund Pädagogik - Sozialwissenschaften - Psychologie (infoconnex)“. Seine Arbeits- und Forschungsschwerpunkte liegen in der Entwicklung internetbasierte Informationssysteme, vor allem in der Nutzung geo-kodierter Informationen bei der Recherche und Ergebnisvisualisierung.