

Zur Leistungsfähigkeit von Crosskordanzen als Mittel der Heterogenitätsbehandlung in Informationssystemen

am Beispiel des
wissenschaftlichen Fachportals infoconnex

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

11. IuK-Jahrestagung 2005

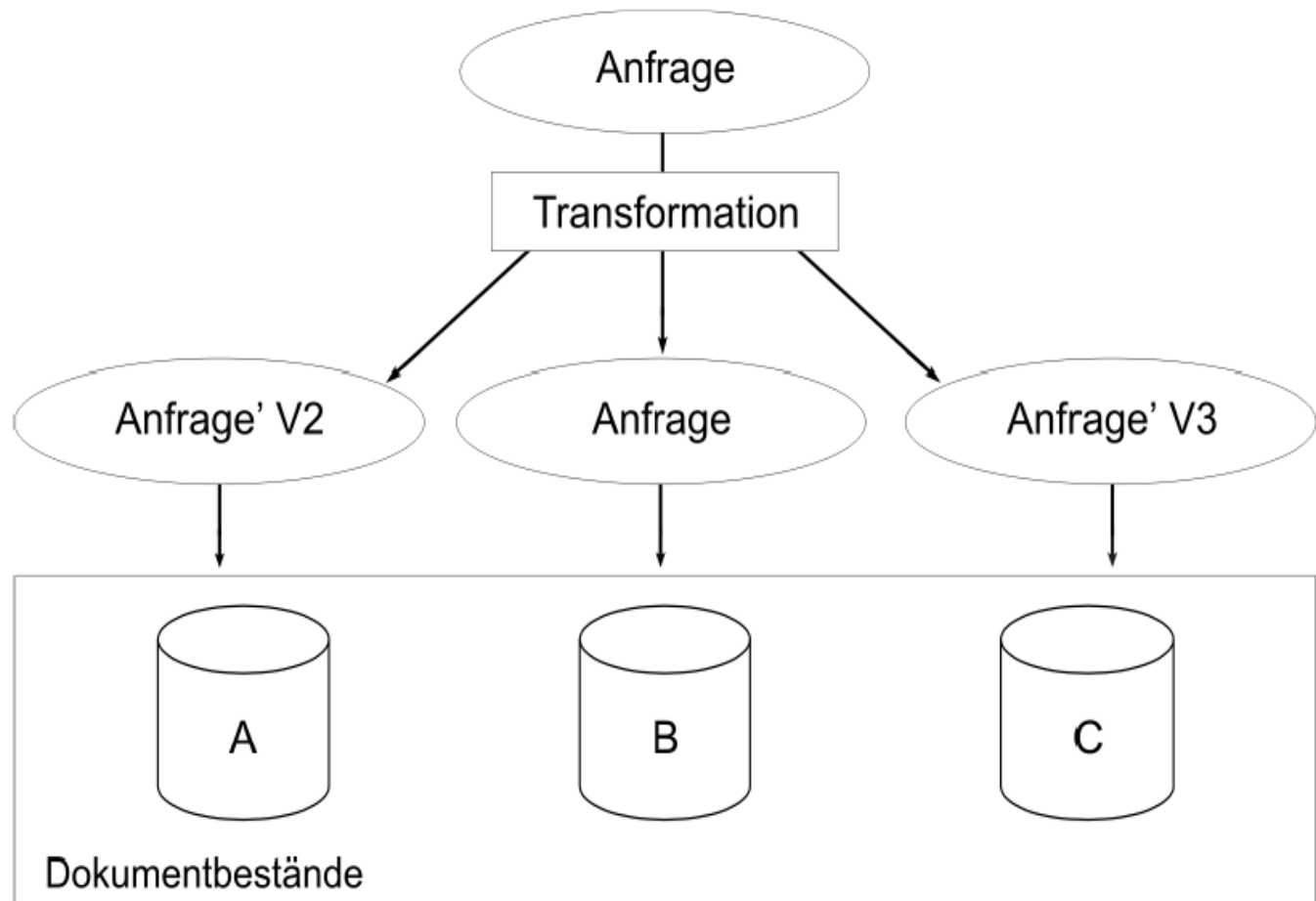


Agenda

- Einführung
 - Heterogenität in Informationssystemen
 - Crosskonkordanzen als Mittel zur Heterogenitätsbehandlung
- infoconnex
 - Die verwendeten Vokabulare
 - Crosskonkordanzen in infoconnex
- Auswertung
 - Vokabularerweiterung durch Crosskonkordanzen
 - Treffererhöhungen für ausgewählte Bereiche des Thesaurus Sozialwissenschaften
- Ausblick

- Fachportale bieten integrierte Recherche über mehrere Informationsangebote
 - Vielzahl unterschiedlicher Informationstypen
 - Unterschiedliche Erschließungssysteme der Informationsanbieter
- negative Auswirkungen auf die Recherche aufgrund der semantischen Heterogenität der Informationsangebote

Heterogenitätsbehandlung: Anfragetransformation





Heterogenitätsbehandlung

Verschiedene Ansätze zur Anfragetransformation,
u.a.:

- Statistische Verfahren
- Intellektuelle Verfahren: **Crosskonkordanzen**

Crosskonkordanzen:

*Gerichtete, relevanzbewertete Relationen
zwischen Termen zweier Thesauri oder
Klassifikationen*

Crosskonkordanzen – Relationsarten

Relationen:

- Äquivalenzrelation
 - sozialer Konflikt = Sozialer Konflikt
 - Legislaturperiode = Wahlperiode
- Oberbegriffsrelation
 - Hochadel < Adel
- Unterbegriffsrelation
 - Kampagne > Werbekampagne
- Ähnlichkeitsrelation
 - Konfliktverhalten \wedge Konfliktfähigkeit

Relevanz (bezogen auf die Treffermenge):

- Hoch
- Mittel
- Gering

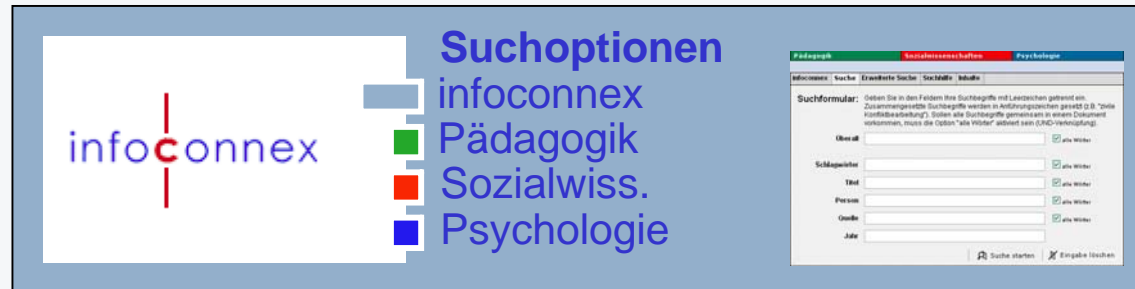
1:1- oder 1:n- Verknüpfungen



Das wissenschaftliche Fachportal infoconnex

- Fachangebote
 - Pädagogik (FIS-Bildung):
ca. 500.000 Nachweise
 - Sozialwissenschaften (SOLIS):
ca. 300.000 Nachweise
 - Psychologie (PSYINDEX):
ca. 185.000 Nachweise
- Fachübergreifende oder fachspezifische Suche

Übersicht infoconnex



Suchoptionen

- infoconnex
- Pädagogik
- Sozialwiss.
- Psychologie

Suchformular: Geben Sie in den Feldern Ihre Suchbegriffe mit Leerzeichen getrennt ein. Zusammengeführte Suchbegriffe werden in Anführungszeichen gesetzt (z. B. "Tolle Korbflügelmaschine"). Sind die Suchbegriffe gemeinsam in einem Dokument vorhanden, muss die Option "als Wörter" aktiviert sein (UND-Verknüpfung).

Überall als Wörter

Schlagwörter als Wörter

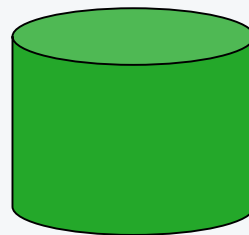
Titel als Wörter

Person als Wörter

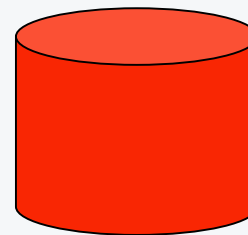
Quelle als Wörter

Jahr als Wörter

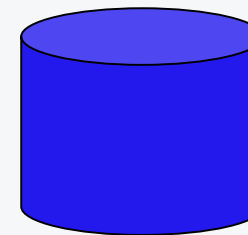
Suche starten Eingabe löschen



FIS-Bildung



SOLIS



PSYINDEX

Thesauri der Datenbanken in infoconnex

- FIS Bildung: Thesaurus Bildung (TheBild)
- SOLIS: Thesaurus Sozialwissenschaften (TheSoz)
- Psyndex: Thesaurus Psychologie (Psyndex Terms – PsyT)

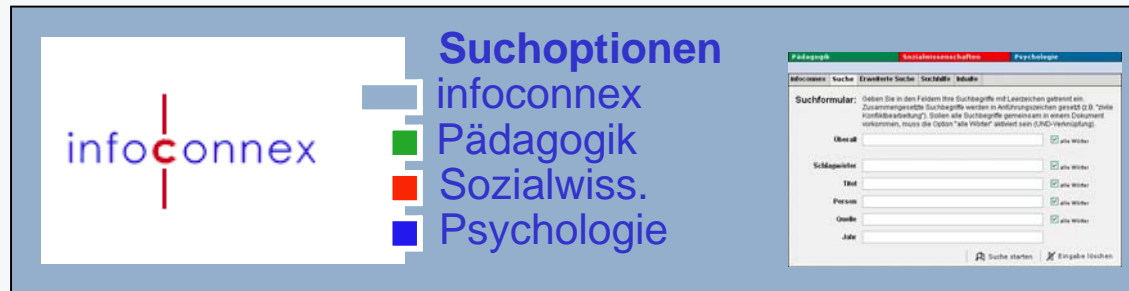
Überschneidungsbereiche (zeichengleiche Terme):

	TheSoz	PsyT	TheBild
TheSoz	7568 (1,00)	763 (0,15)	5520 (0,10)
PsyT	763 (0,10)	4970 (1,00)	1390 (0,03)
TheBild	5520 (0,73)	1390 (0,28)	52643 (1,00)

Beispiel:

Der TheSoz und die PsyT haben 763 zeichengleiche Deskriptoren, d.h. 10% des TheSoz sind in den PsyT enthalten.

Fachübergreifende Suche



Suchoptionen

- infoconnex
- Pädagogik
- Sozialwiss.
- Psychologie

Suchformular: Geben Sie in den Feldern Ihre Suchbegriffe und Lesarten ein. Zusammengeführte Suchbegriffe werden in nachfolgenden queries (z.B. "Title AND Abstract"), sollen alle Suchbegriffe gemeinsam in einem Dokument vorkommen, muss die Option "alle Wörter" aktiviert sein (AND-Verknüpfung).

Übersicht: alle Wörter

Schlagwörter: alle Wörter

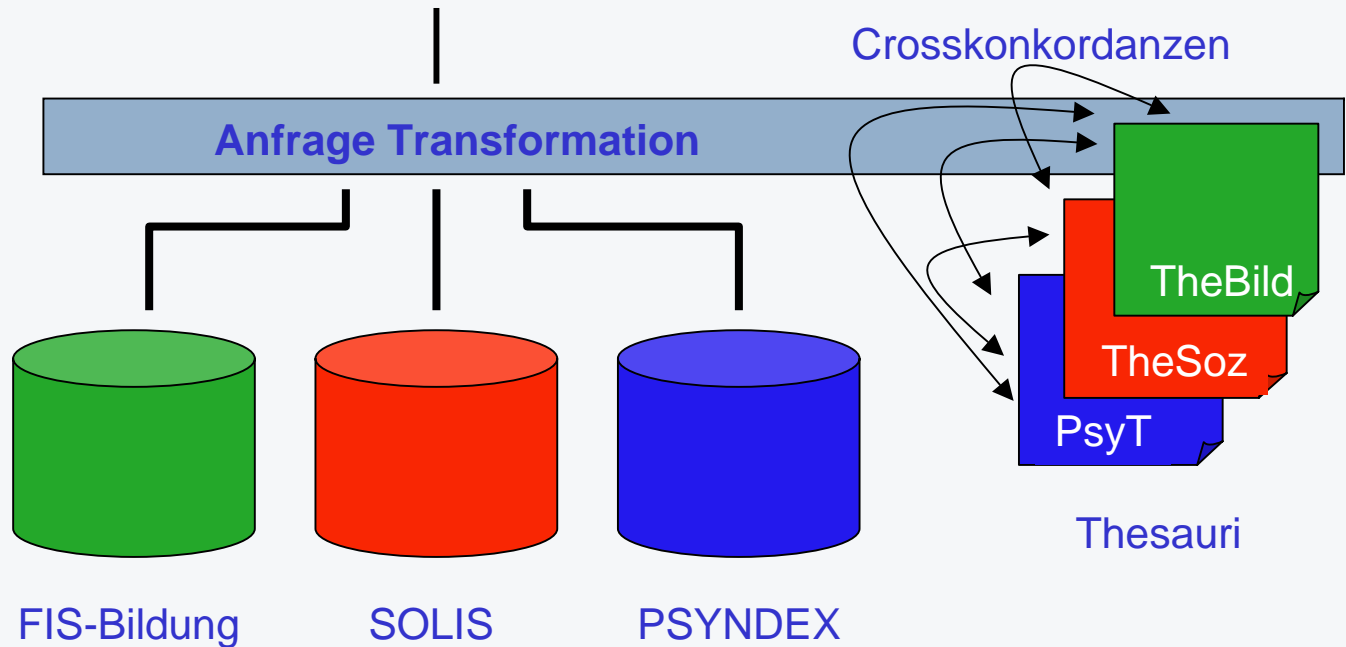
Titel: alle Wörter

Person: alle Wörter

Quelle: alle Wörter

Jahr: alle Wörter

Suche starten Eingabe löschen



infoconnex – Beispiel-Transformationen

- infoconnex-Suche nach „Freizeitpädagogik“
- Konkordanz TheSoz → PsyT:
Freizeitpädagogik =+ Bildung und Erziehung + Freizeit
- Ausschnitt des Ergebnisfensters:

Sozialwissenschaften

Suchformular > **Trefferliste**

Anfrage

Schlagwort=freizeitpaedagogik[97]

Psychologie

Suchformular > **Trefferliste**

Anfrage

Schlagwort=freizeitpädagogik -> (bildung und erziehung und freizeit)

Aktuell nur Verwendung der Äquivalenzrelationen.

Untersuchte Bereiche:

- 1) Vokabularerweiterung ausgehend von einem Thesaurus
- 2) Treffererhöhungen durch Crosskonkordanzen
 - für ausgewählte Systematikstellen des ThesSoz
 - für die am häufigsten und die am wenigsten verwendeten Deskriptoren des TheSoz

Auswertung – Vokabularerweiterung

Richtung	Desk.	Äquiv.	Oberbegr.	Unterbegr.	Ähnl.	gesamt
TheSoz->PsyT	7568	584 (7,7%)	1084 (13,9%)	1048 (13,9%)	1628 (21,5%)	57,4%
TheSoz->TheBild	7568	854 (11,3%)	210 (2,8%)	46 (0,6%)	142 (1,9%)	16,5%
PsyT->TheSoz	4970	301 (6,0%)	1451 (29,2%)	37 (0,7%)	1843 (37,1%)	73,1%
PsyT->TheBild	4970	1769 (35,6%)	1075 (21,6%)	18 (0,4%)	1680 (33,8%)	91,4%
TheBild->TheSoz	52643	551 (1,0%)	3 (0,0%)	45 (0,1%)	88 (0,2%)	1,3%
TheBild->PsyT	52643	554 (1,0%)	994 (1,9%)	379 (0,7%)	832 (1,6%)	5,3%

Auswertung – Vokabularerweiterung

- Auswirkungen der Größe des Thesaurus Bildung
- viele Ähnlichkeitsrelationen in den thematisch nahen Thesauri TheSoz und PsyT
- Vokabularerweiterung pro Konkordanz: durchschnittlich 27%
- Vokabularerweiterung pro Thesaurus:

TheSoz	68,70%
PsyT	87,48%
TheBild	6,59%
Durchschnitt	54,26%

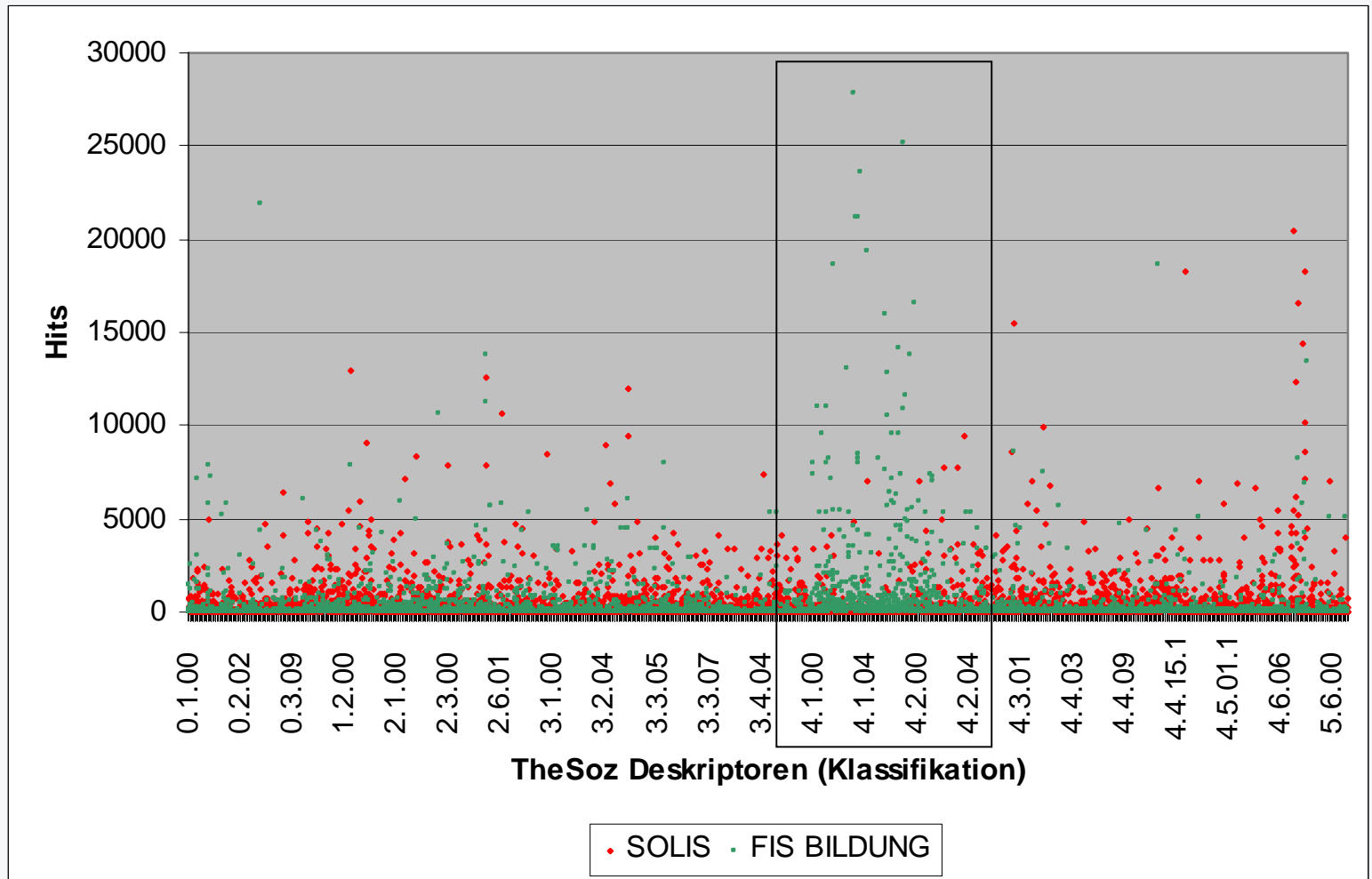


Übersicht Treffererhöhungen

Treffererhöhungen durch Äquivalenzrelationen,
ausgehend vom Vokabular des TheSoz

1. Untersuchung pro Konkordanz, dabei Spezialisierung auf bestimmte Klassifikationsstellen
 - TheSoz → TheBild
 - TheSoz → PsyT
2. Untersuchung der am häufigsten und der am wenigstens verwendeten Deskriptoren des TheSoz

Treffererhöhungen TheSoz → TheBild: Gesamte Klassifikation

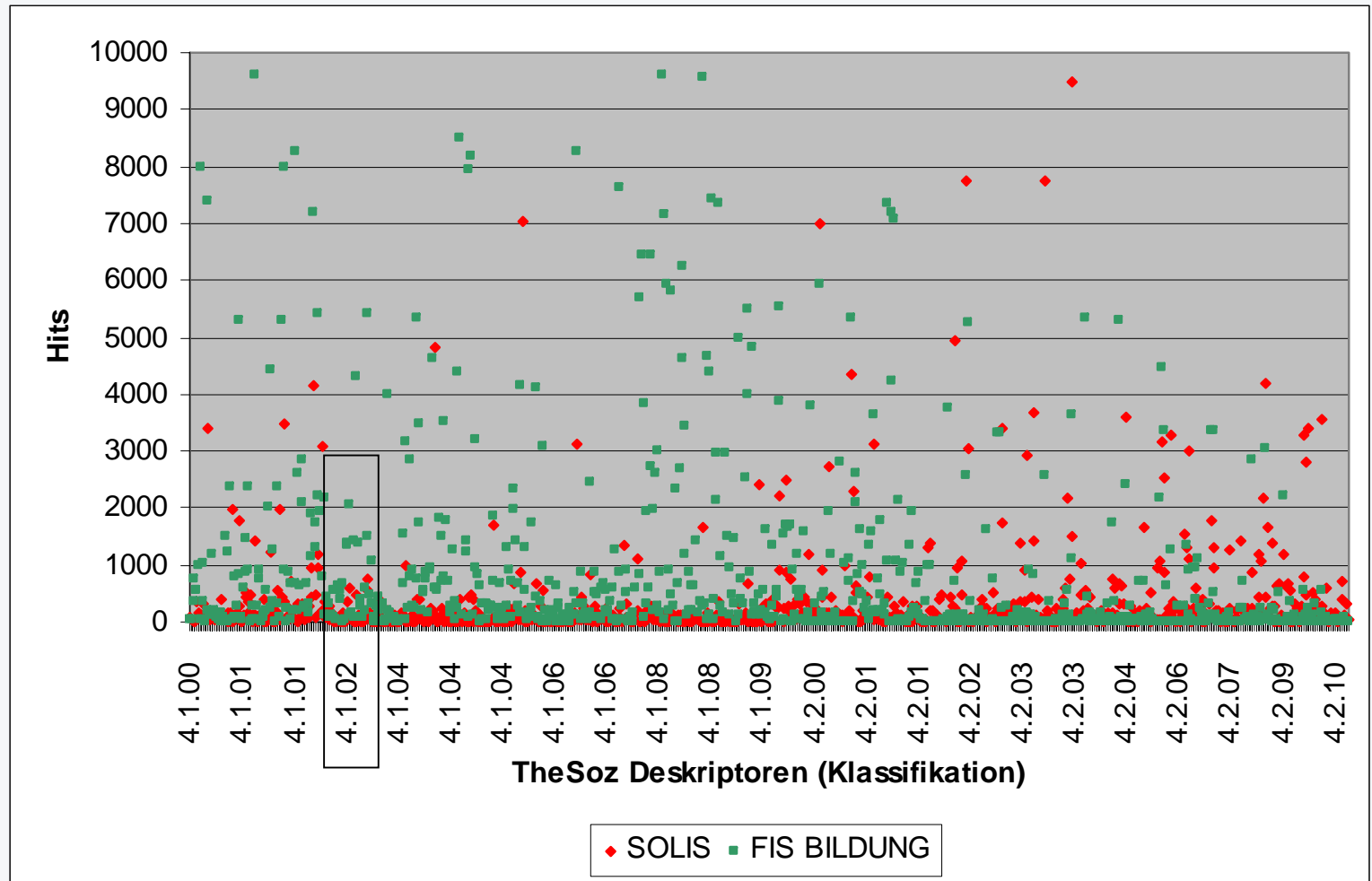


Trefferhäufigkeit im Klassifikationsbereich 4.1-4.2

4.1 Erziehung & Bildung

4.2 Kommunikation, Sprache, Publizistik

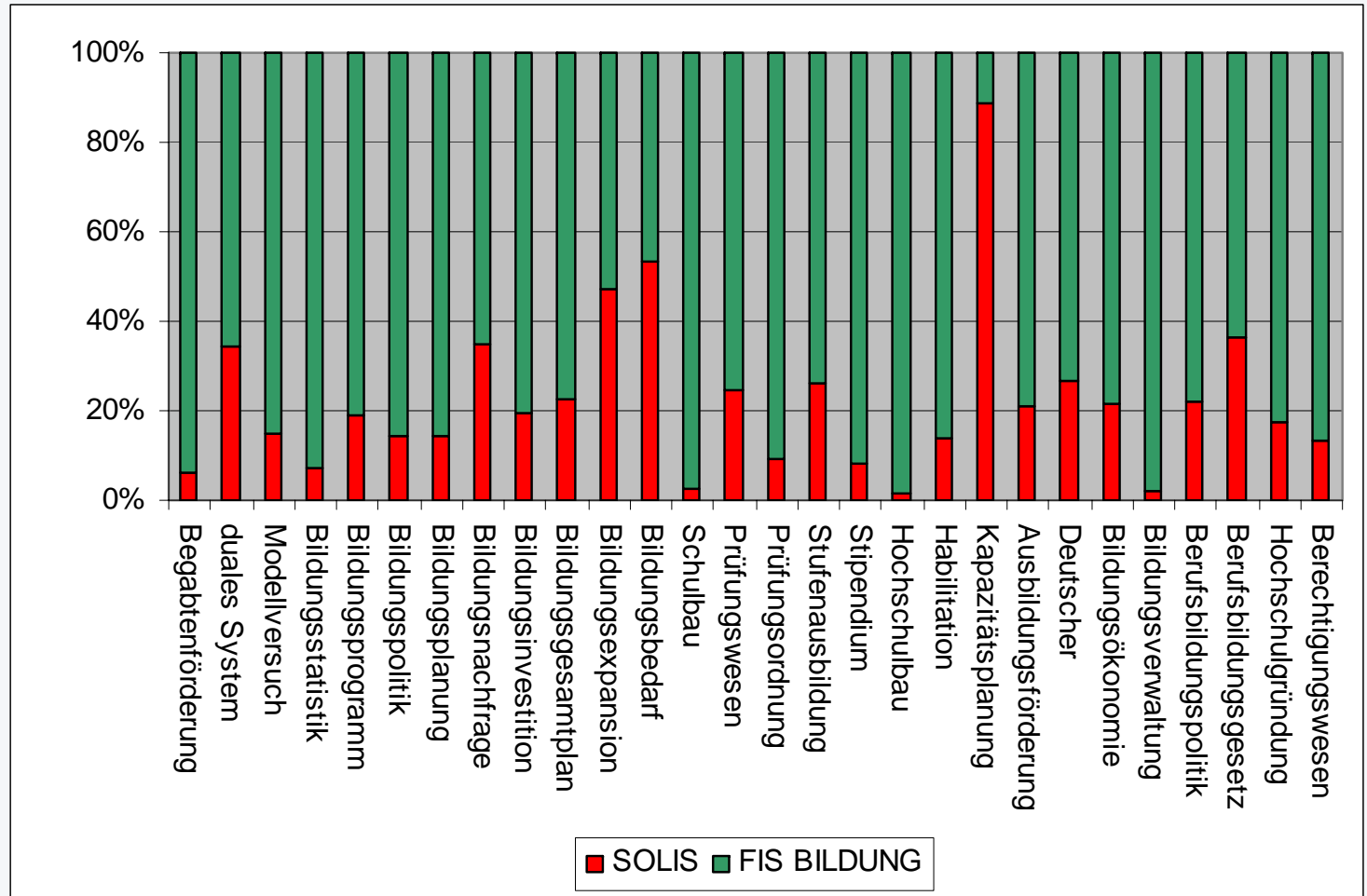
Treffererhöhungen TheSoz → TheBild: Klassifikationsstellen 4.1 - 4.2



4.1 Erziehung & Bildung

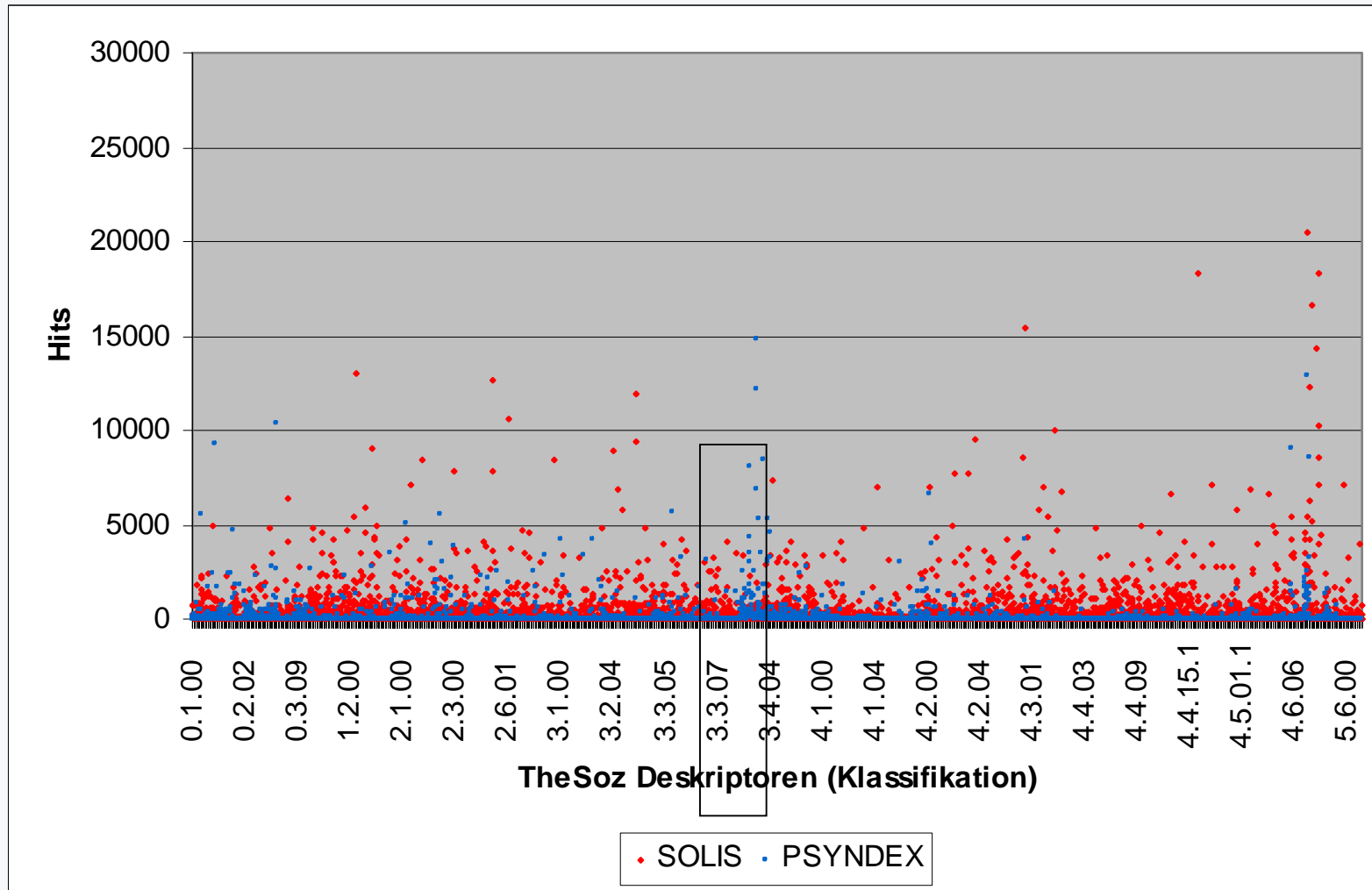
4.2 Kommunikation, Sprache, Publizistik

Treffererhöhungen TheSoz → TheBild: Klassifikationsstelle 4.1.02



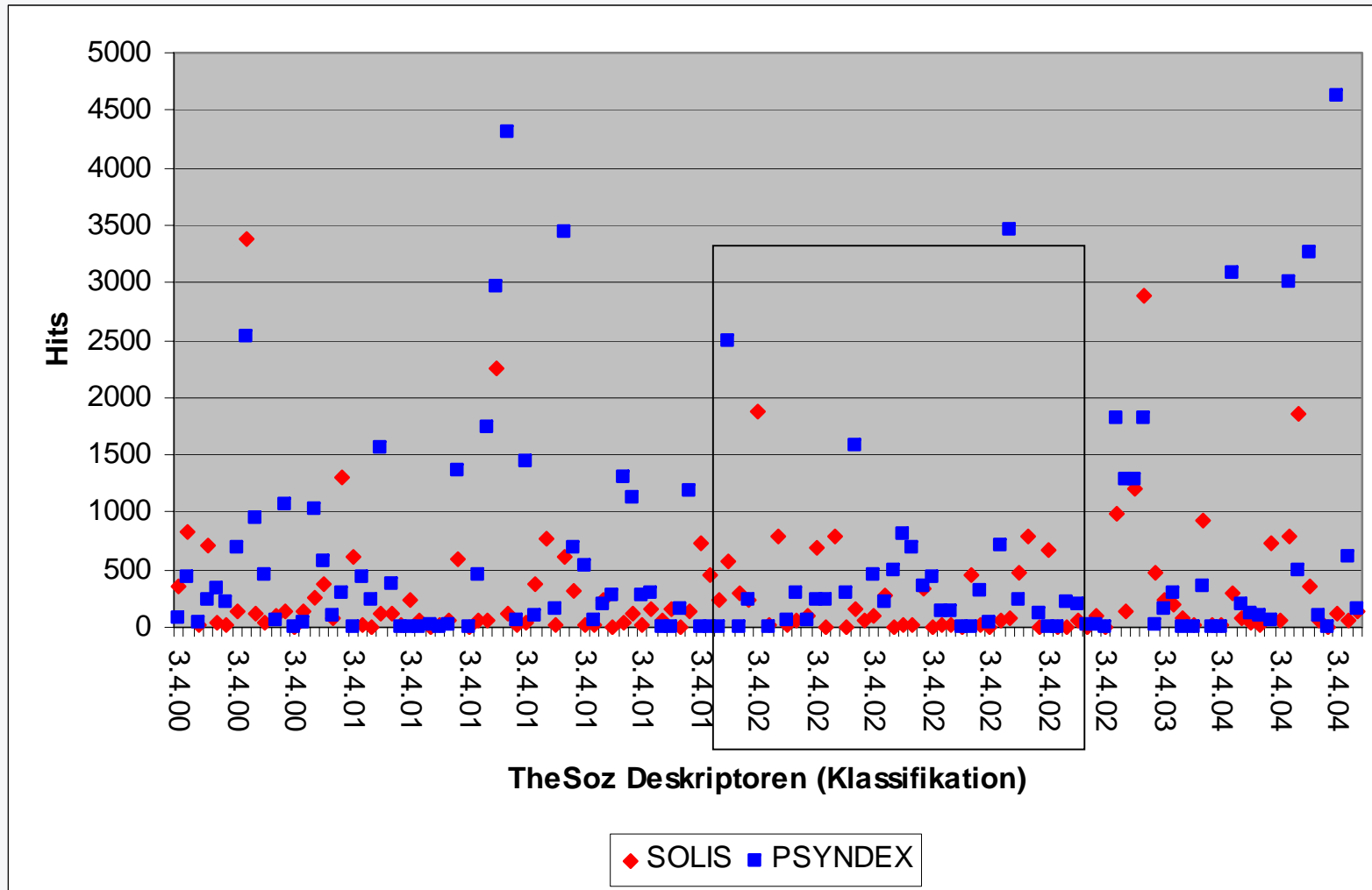
4.1.02 Bildungsplanung, Bildungspolitik, Bildungsökonomie

Treffererhöhungen TheSoz → PsyT: Gesamte Klassifikation



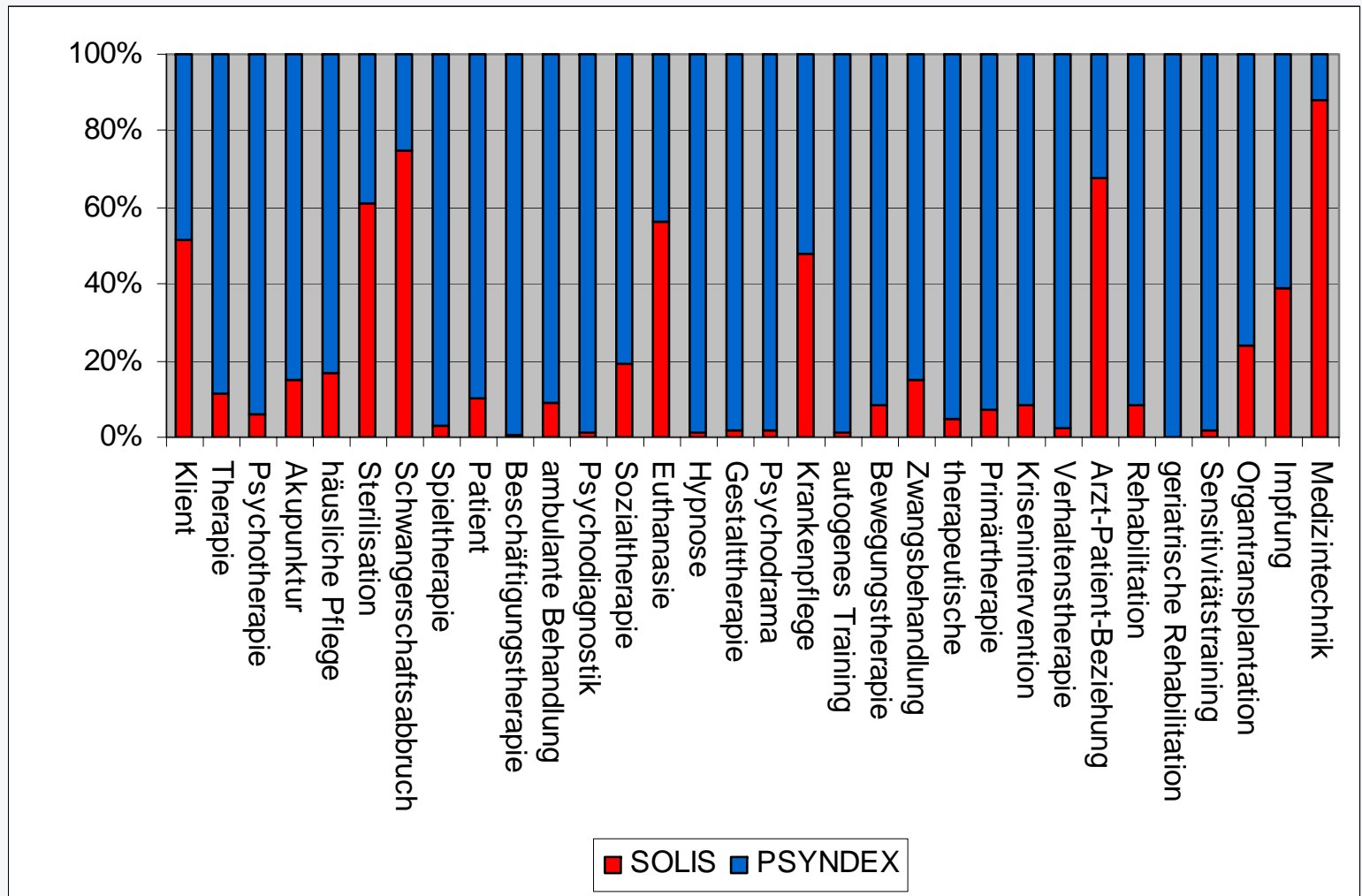
3.4 Medizin, Gesundheitswesen

Treffererhöhungen TheSoz → PsyT: Klassifikationsstelle 3.4



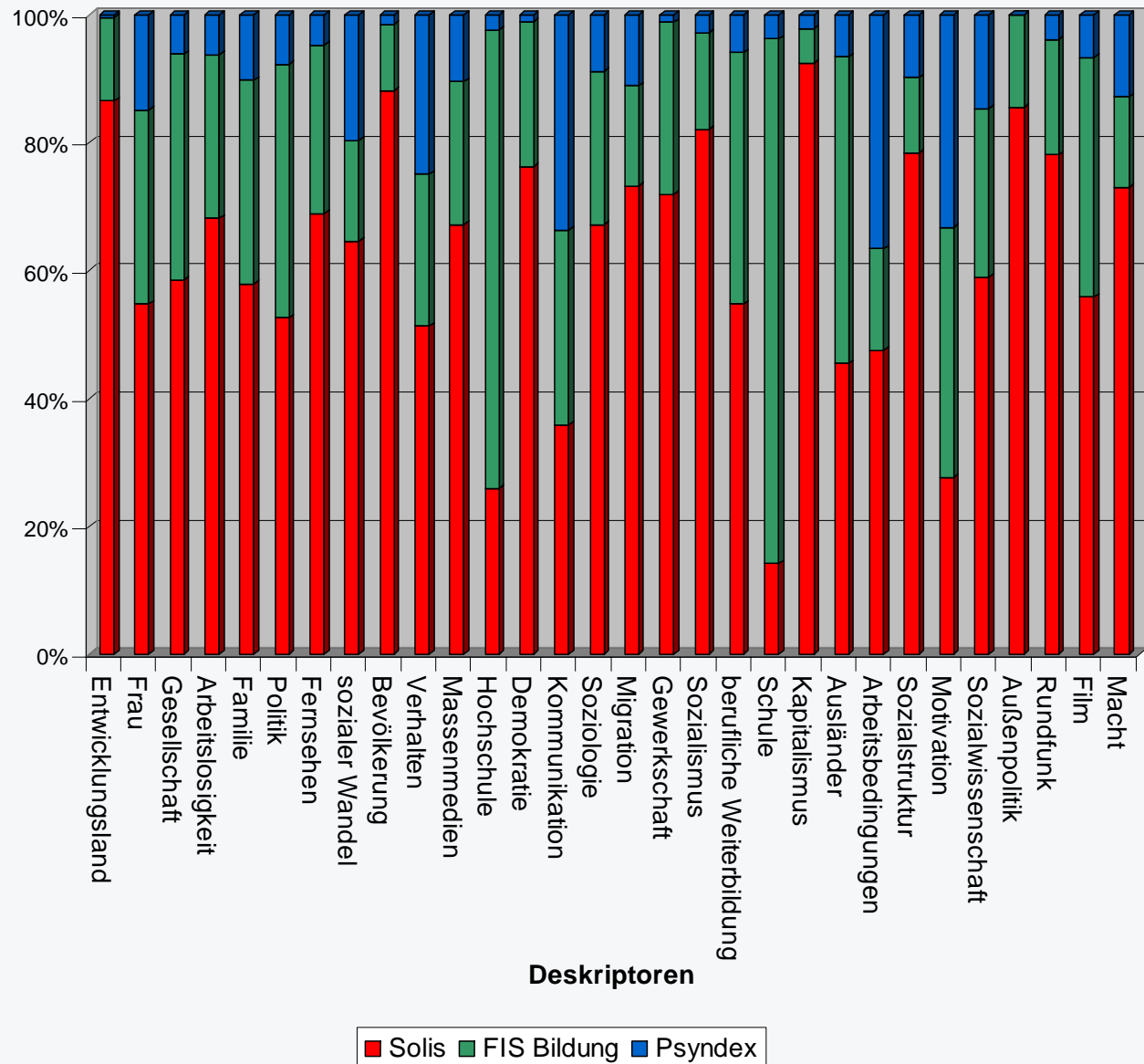
3.4 Medizin, Gesundheitswesen

Treffererhöhungen TheSoz → PsyT: Klassifikationsstelle 3.4.02

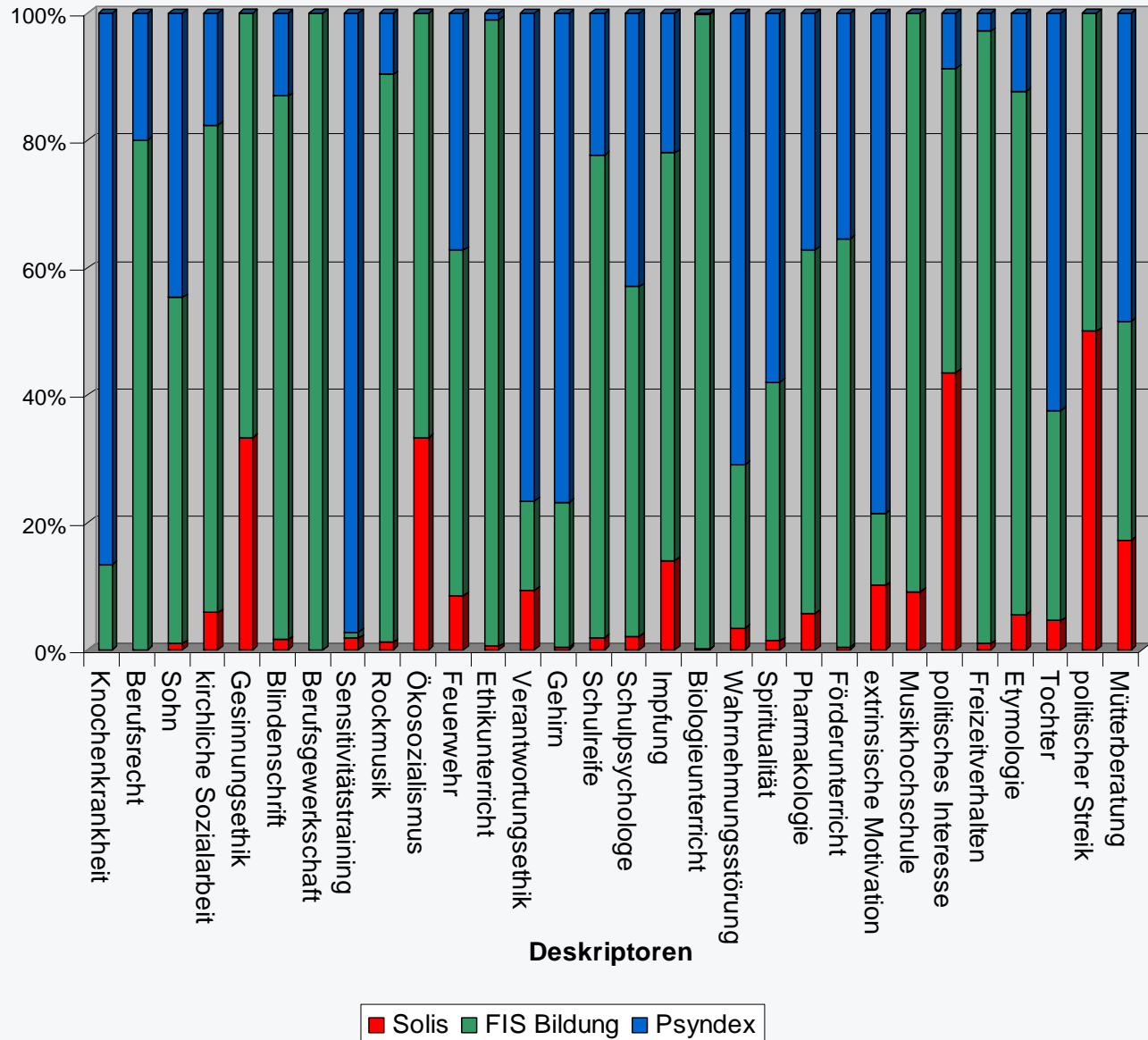


3.4.02 Prävention, Therapie, medizinische Rehabilitation

30 TheSoz-Deskriptoren, mit denen am häufigsten indexiert wird



30 TheSoz-Deskriptoren, mit denen am wenigsten indexiert wird





Zusammenfassung

- signifikante Vokabularerweiterung (im Schnitt ca. 50% pro Fachgebiet)
- Treffererhöhung kann als nachgewiesen gelten, besonders in interdisziplinären Bereichen.
- Vereinfachter Zugang (1 Vokabular)
- noch keine Aussagen zur Relevanz der zusätzlichen Treffer

- weiterer Einsatz von Crosskonkordanzen für das IR
 - Innerhalb eines Fachgebiets (DZI, THESOZ)
 - Zwischen verwandten Fachgebieten
 - Indirekte Term-Transfers
- Modellierung des Einsatzes der bislang nicht verwendeten Relationen (\wedge , $>$, $<$)
- Recall- und Precision-Untersuchungen



Vielen Dank für Ihre Aufmerksamkeit!

Anne-Kathrin Walter, Philipp Mayr,
Maximilian Stempfhuber, Arne Ballay
{walter,mayr,stempfhuber,ballay}@bonn.iz-soz.de

<http://www.gesis.org/IZ/>