

# *Auf dem Weg zum wissenschaftlichen Fachportal – Modellbildung und Integration heterogener Informationssammlungen*

Philipp Mayr, Maximilian Stempfhuber & Anne-Kathrin Walter  
Informationszentrum Sozialwissenschaften (IZ)

Frankfurt am Main, den 23. Mai 2005  
27. Online-Tagung der DGI 2005



## Agenda

1. Ausgangssituation
2. Modellbildung & Heterogenitätsbehandlung
  - Kompetenzzentrum
  - Modellbildung als Voraussetzung für Integration
  - Konzept: Wissenschaftliches Fachportal
  - Heterogenitätsbehandlung mittels Crosskonkordanzen (CK)
3. Exkurs Treffererhöhung durch CK
  - Beispiele (infoconnex)
  - Effekte
4. Ausblick



## Ausgangssituation

- Polyzentrische Informationsversorgung
- Widersprüchliche Nutzeranforderung (z.B. schnell vs. umfassend)
- Trennung in kommerzielle und freizugängliche Materialien
- geändertes Informationsverhalten, siehe aktuelle Nutzerbefragungen/Studien\*
  - Bündelung (Informationstypen)
  - Integrierte Suche
  - Direkter Zugriff (Volltext)
  - Interdisziplinärität (Fachcluster)
  - Relevante, qualitative Dokumente

\* Quellen:

Boekhorst et al. 2003, Poll 2004, IMAC 2002, RSLG 2002, Binder et al. 2001, Stahl et al. 1998, zu WWW-Suchmaschinen: Machill & Welp 2003



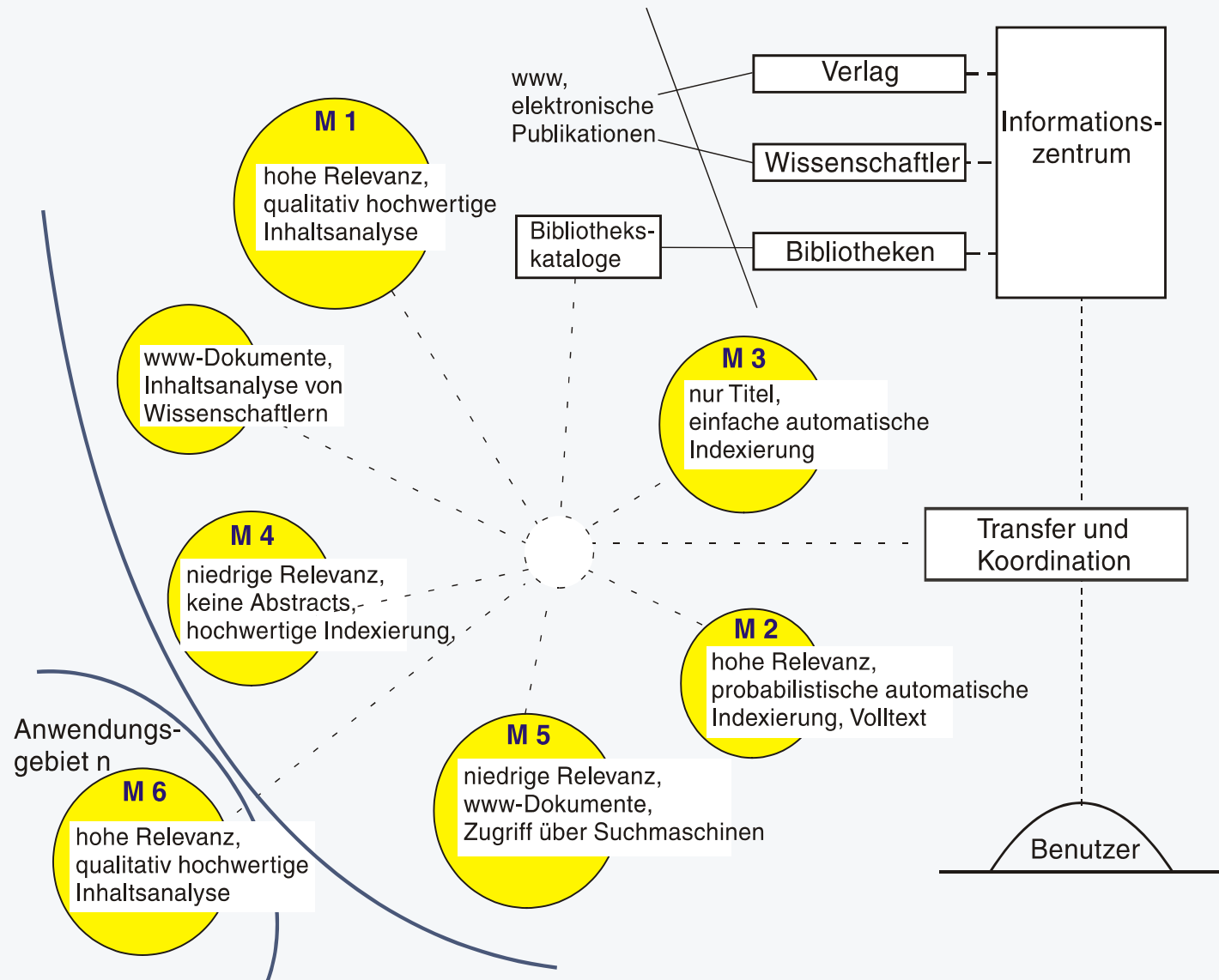
## Aktuelle Entwicklungen

1. Open Access als alternative Publikationsform
  - steigende Zahl und Akzeptanz OA-Zeitschriften und Eprints-Archiven (Harnad et al. 2004)
  - ➔ Tendenz zu mehr Volltexten (self-archiving)
2. Anwendung von Suchmaschinen-Technologie auf wissenschaftlichen Dokumentenraum
  - Beispiele in Deutschland
    - > FAST-Implementationen: BASE, hbz
  - Google Scholar
  - Überlegung: vascoda mit FAST-Technologie (z.B. Schöning-Walter auf IuK 2005)



# Polyzentrischer Dokumentenraum

J. Krause





## Kompetenzzentrum

Teilprojekt innerhalb des Kompetenznetzwerks „Neue Dienste, Standardisierung, Metadaten“ (BMBF)

- Modellbildung: übergreifende Modellbildung für komplexe Informationsinfrastrukturen, u. a. am Beispiel des Wissenschaftsportals vascoda mit allen nachgeschalteten Ebenen
  - Heterogenitätsbehandlung: als notwendige Ergänzung zur Standardisierung durch einheitliche Metadaten
- ➔ Vorschläge (Modelle) zur Strukturierung wissenschaftlicher Informationsangebote
- ➔ Schaffung von Rahmenbedingungen und Infrastrukturen zur Förderung der semantischen Integration heterogener Informationssammlungen

Projektlaufzeit September 2004 bis August 2007



## Modellbildung

Ziel: ein informationswiss. basiertes Rahmenkonzept (Modell) zur Kombination bestehender Einzellösungen in eine stimmige Informationsarchitektur

1. Analyse der vorliegenden Fachinformations-Strukturen
2. Entwicklung eines abstrakten Modells für übergreifende, kooperative Angebotsstrukturen

Fragen:

- Kombination/Verteilung komplementärer Informationsarten für spezifische Nutzer
- Ableitung sinnvoller Recherchealgorithmen (Einbeziehung d. Standardisierungskomponenten)



## Wissenschaftliche Fachportale

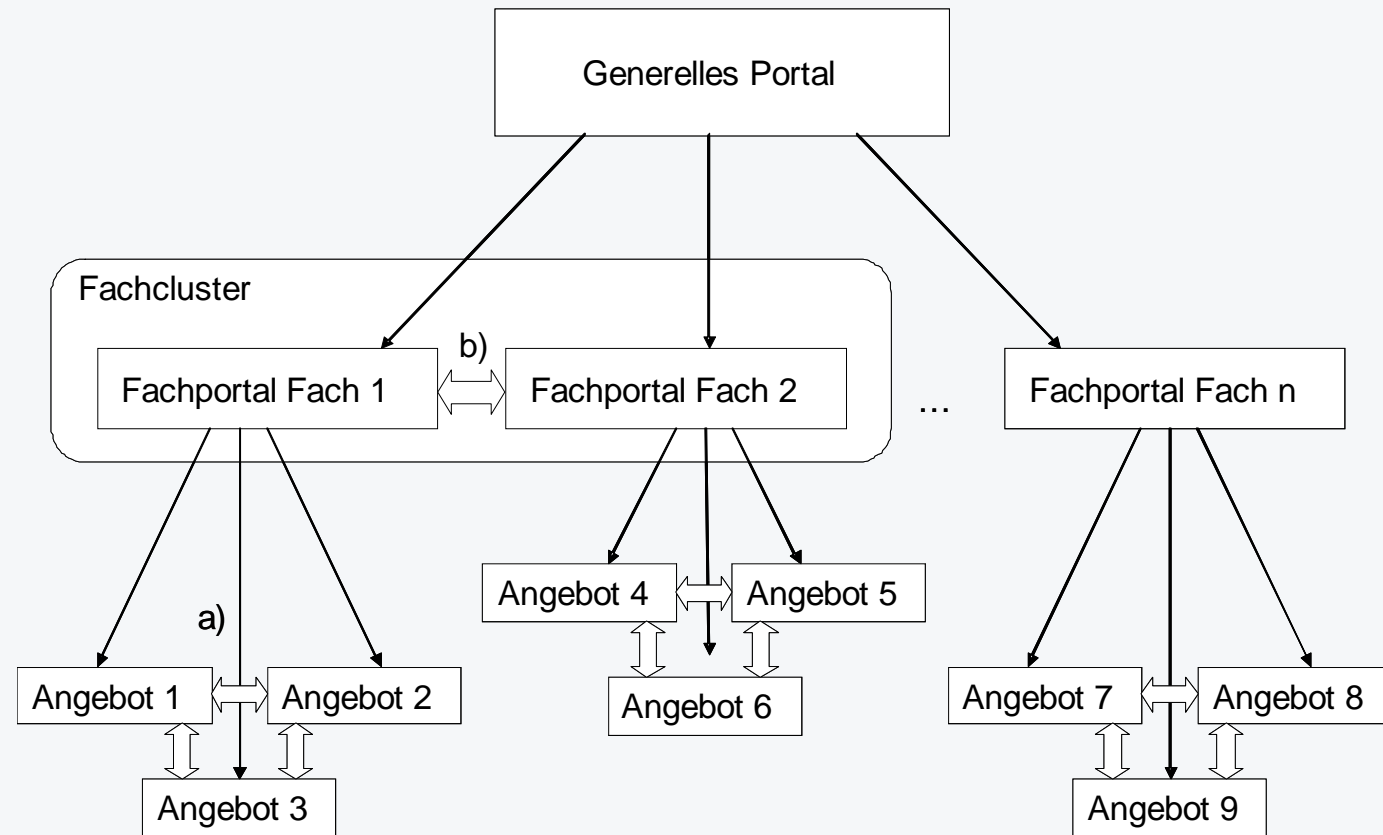
„Wissenschaftliche Fachportale sollen möglichst alle aus Nutzersicht relevanten Informationen eines Faches integrieren, wobei das Augenmerk auf der inhaltlichen und konzeptuellen Integration liegt und die rein technische Integration als notwendige Voraussetzung in den Hintergrund rückt.“

- Baustein innerhalb komplexer, fachübergreifender und dezentraler Portale
- auf einer mittleren Abstraktionsebene zwischen singulären Informationsangeboten und generellen Portalen

Ziel: Integration paralleler Portale innerhalb eines Faches zu einem Fachportal

## Beispiel

### Kaskadierendes Modell einer Portalinfrastruktur





## Wesentliche Anforderungen für wiss. Fachportale

- Integration unterschiedlicher Informationstypen
  - Opacs
  - Fachdatenbanken
  - Volltexte
  - Internetquellen
- Kombination und Integration verschiedener Erschließungsformen (formal & inhaltlich)
- Kopplung der Informationssysteme im Hinblick auf
  - Aktualität
  - Vollständigkeit
- Transparenz der Komplexität für den Nutzer (Benutzerschnittstelle)



## Heterogenitätsbehandlung

Verschiedene Ansätze zur Anfragetransformation als Maßnahmen zur Integration heterogen erschlossener Informationssammlungen

- Statistische Verfahren
- (Deduktive Verfahren)
- Intellektuelle Verfahren: **Crosskonkordanzen**

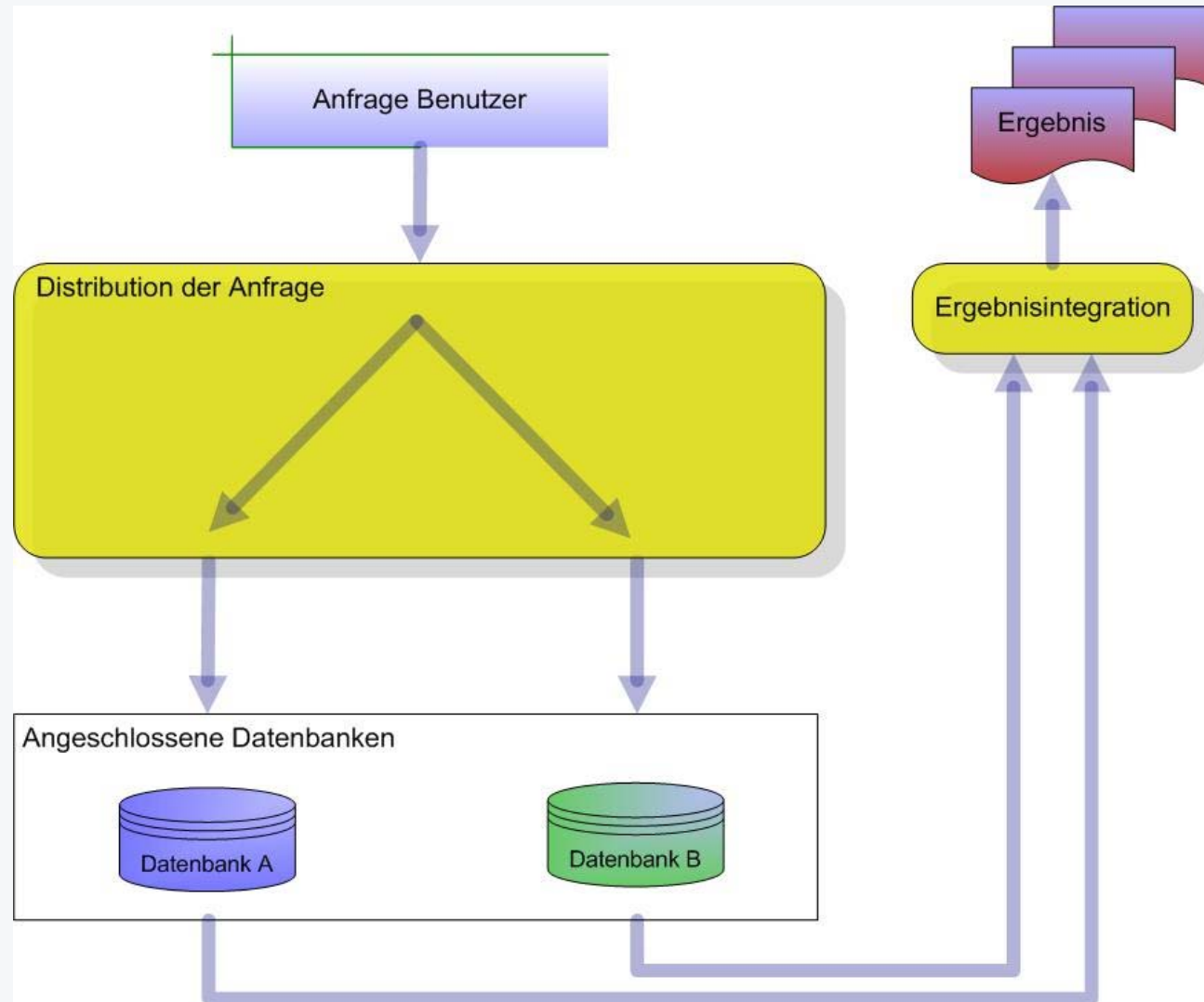
Crosskonkordanzen:

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



# Heterogenitätsbehandlung

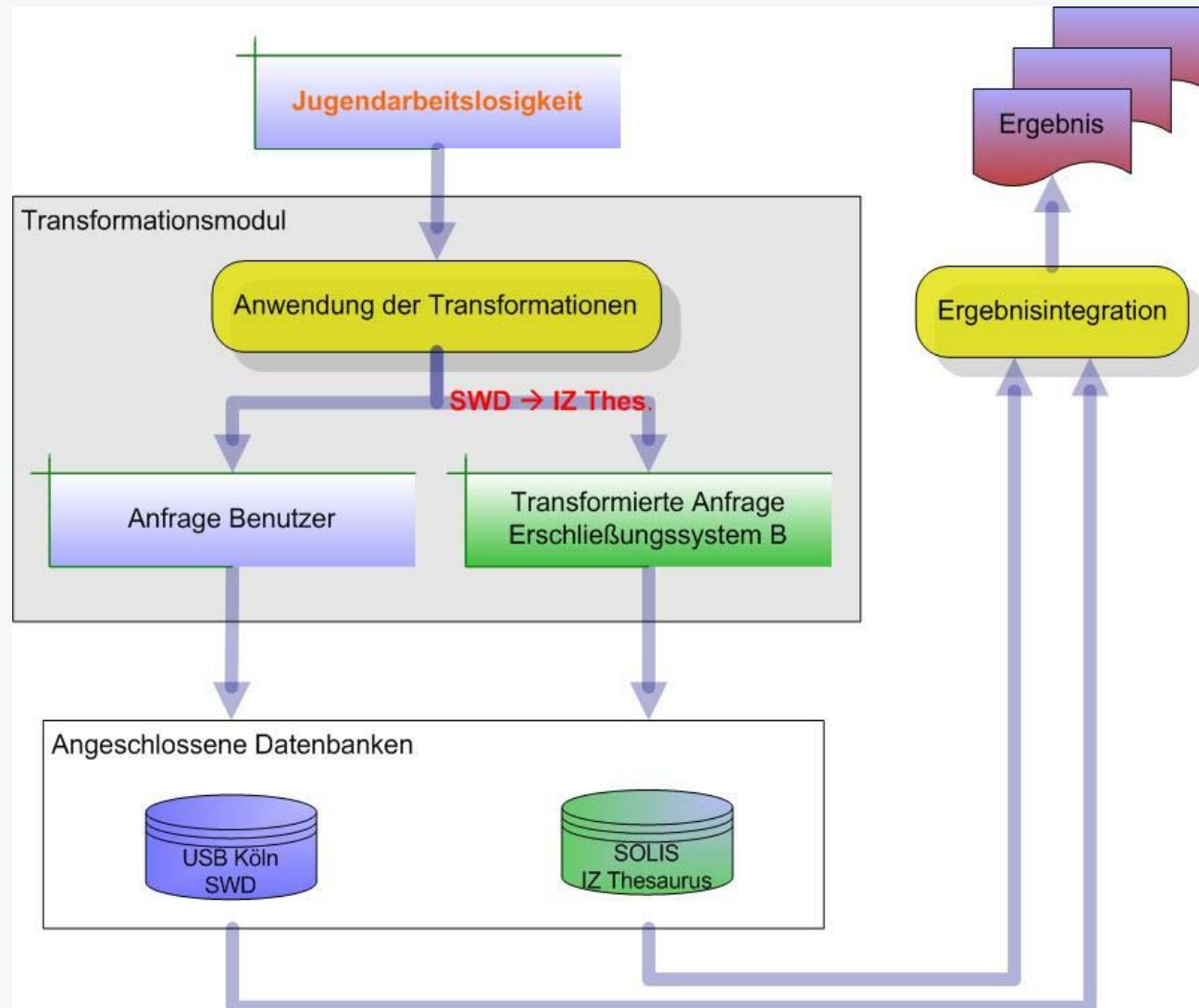
m. müller





# Heterogenitätsbehandlung

m. müller





## Crosskonkordanzen – Relationsarten

### Relationen:

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

### Relevanz (bezogen auf die Treffermenge):

- Hoch
- Mittel
- Gering

1:1- oder 1:n- Verknüpfungen



## infoconnex – Beispiel-Transformationen

- infoconnex-Suche nach „Freizeitpädagogik“
- Konkordanz TheSoz → PsyT:

Freizeitpädagogik =+ Bildung und Erziehung + Freizeit

**Sozialwissenschaften**

Suchformular > **Trefferliste**

**Anfrage**

Schlagwort=freizeitpaedagogik[97]

**Psychologie**

Suchformular > **Trefferliste**

**Anfrage**

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

Aktuell nur Verwendung der Äquivalenzrelationen.



## Exkurs Treffererhöhungen durch CK (infoconnex)

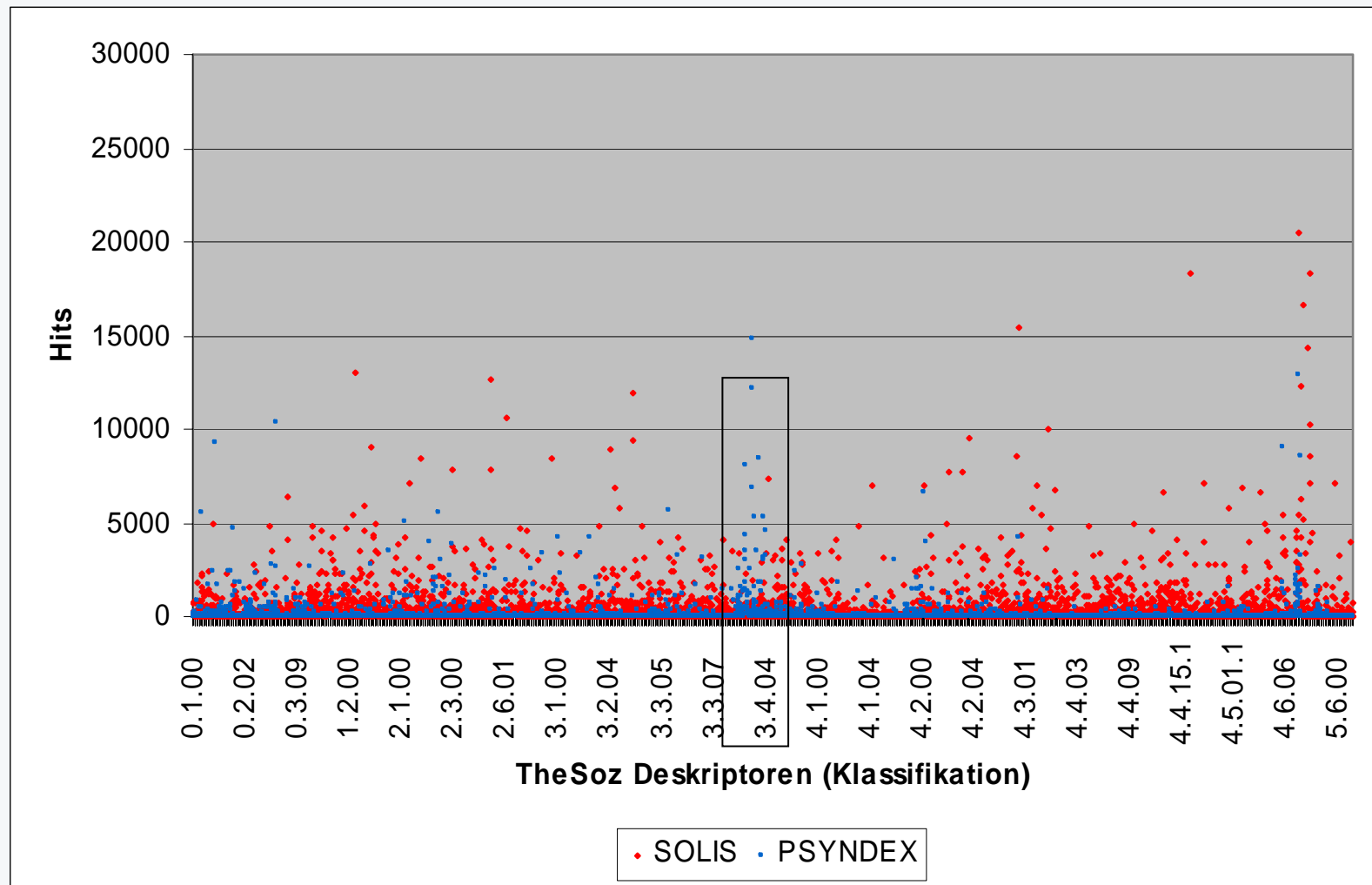
Treffererhöhungen durch Äquivalenzrelationen,  
ausgehend vom Vokabular des TheSoz

TheSoz → PsyT

1. Untersuchung einer bestimmten Klassifikationsstelle
2. Untersuchung der häufigsten und der seltensten geindexten Deskriptoren des TheSoz

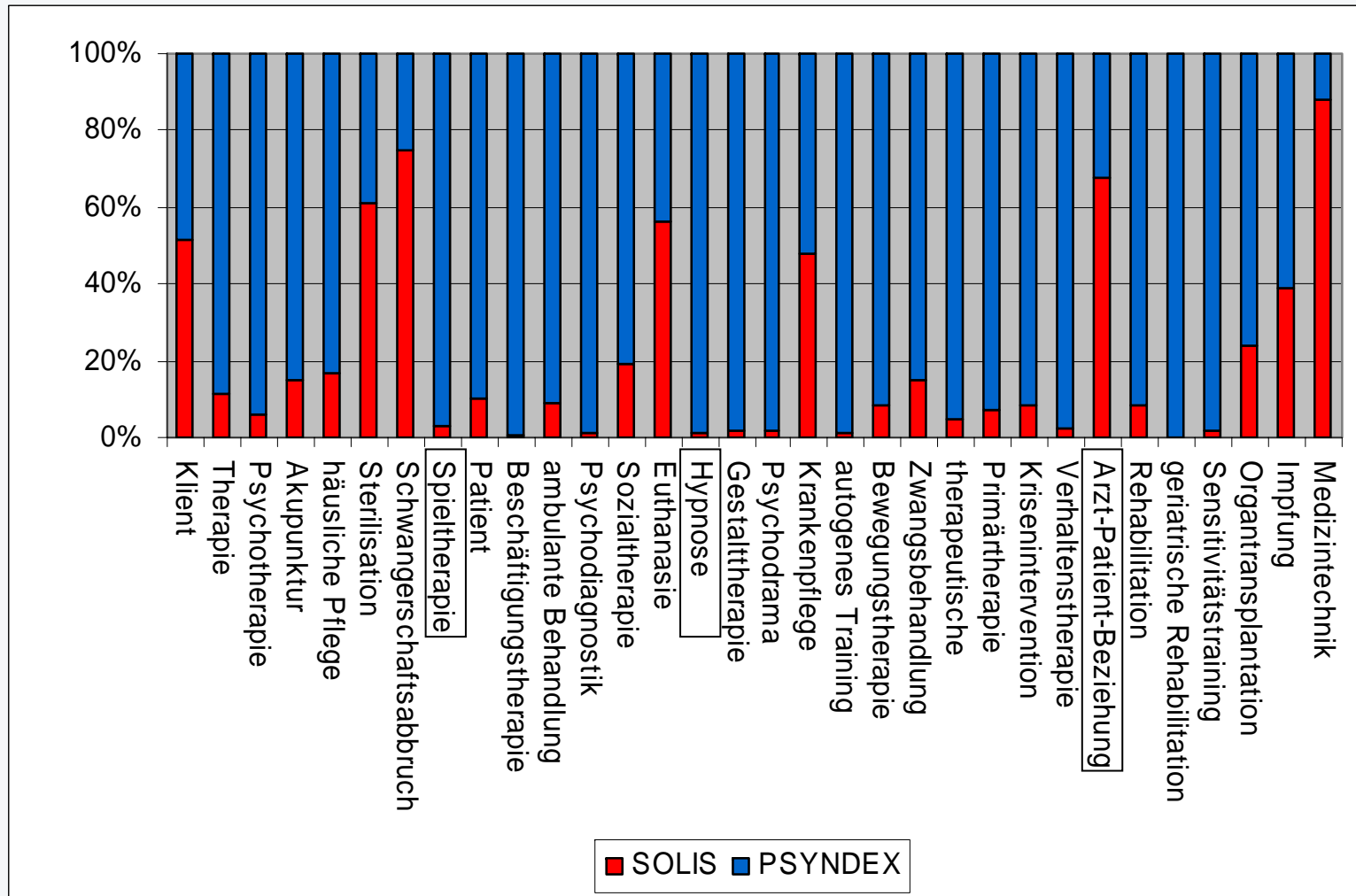
Hinweis: Die folgenden Auswertung stehen erst am Anfang.

# Treffererhöhungen TheSoz → PsyT: Gesamte Klassifikation (TheSoz)



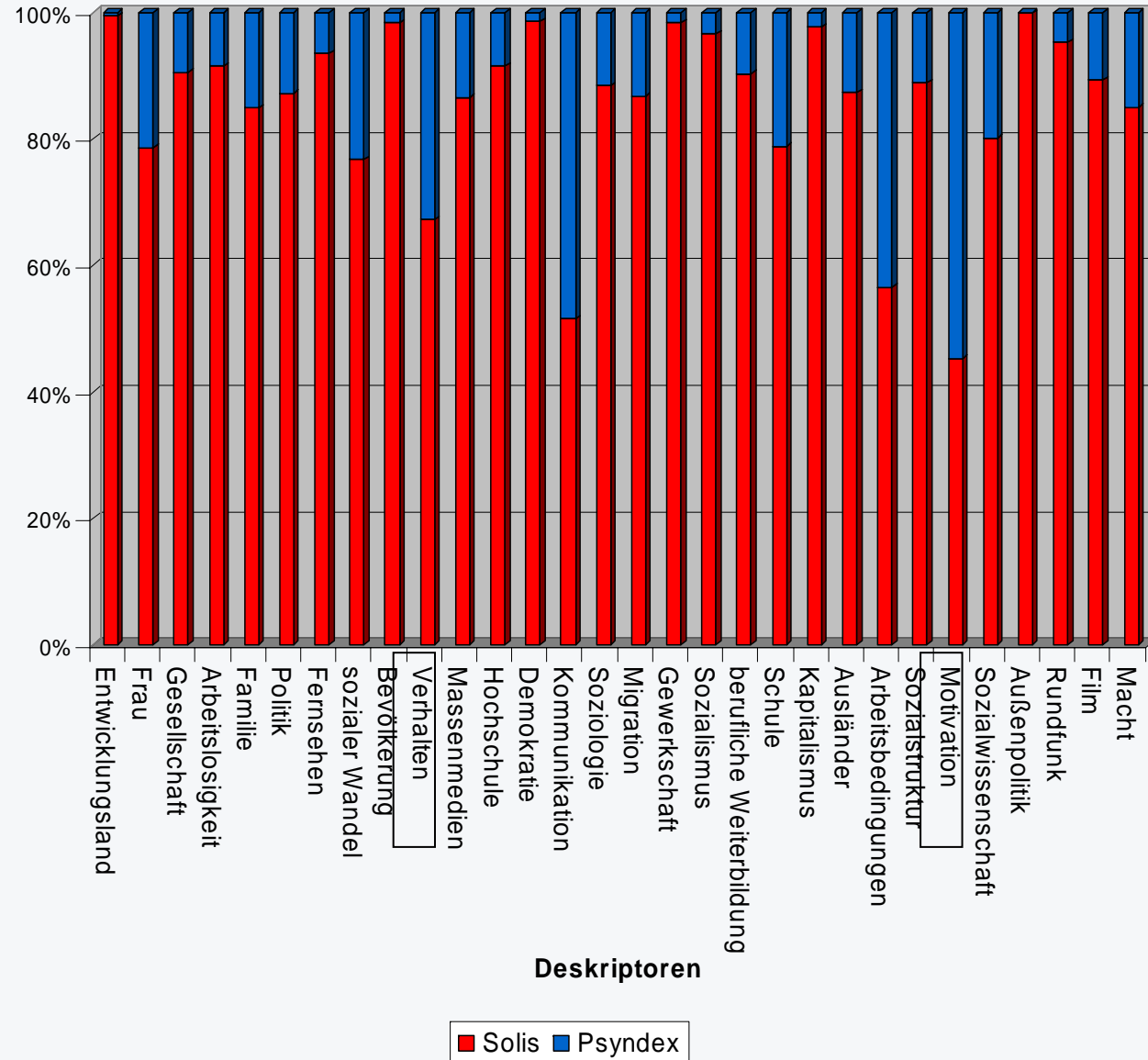
## 3.4 Medizin, Gesundheitswesen

# Treffererhöhungen TheSoz → PsyT: Klassifikationsstelle 3.4.02 (TheSoz)

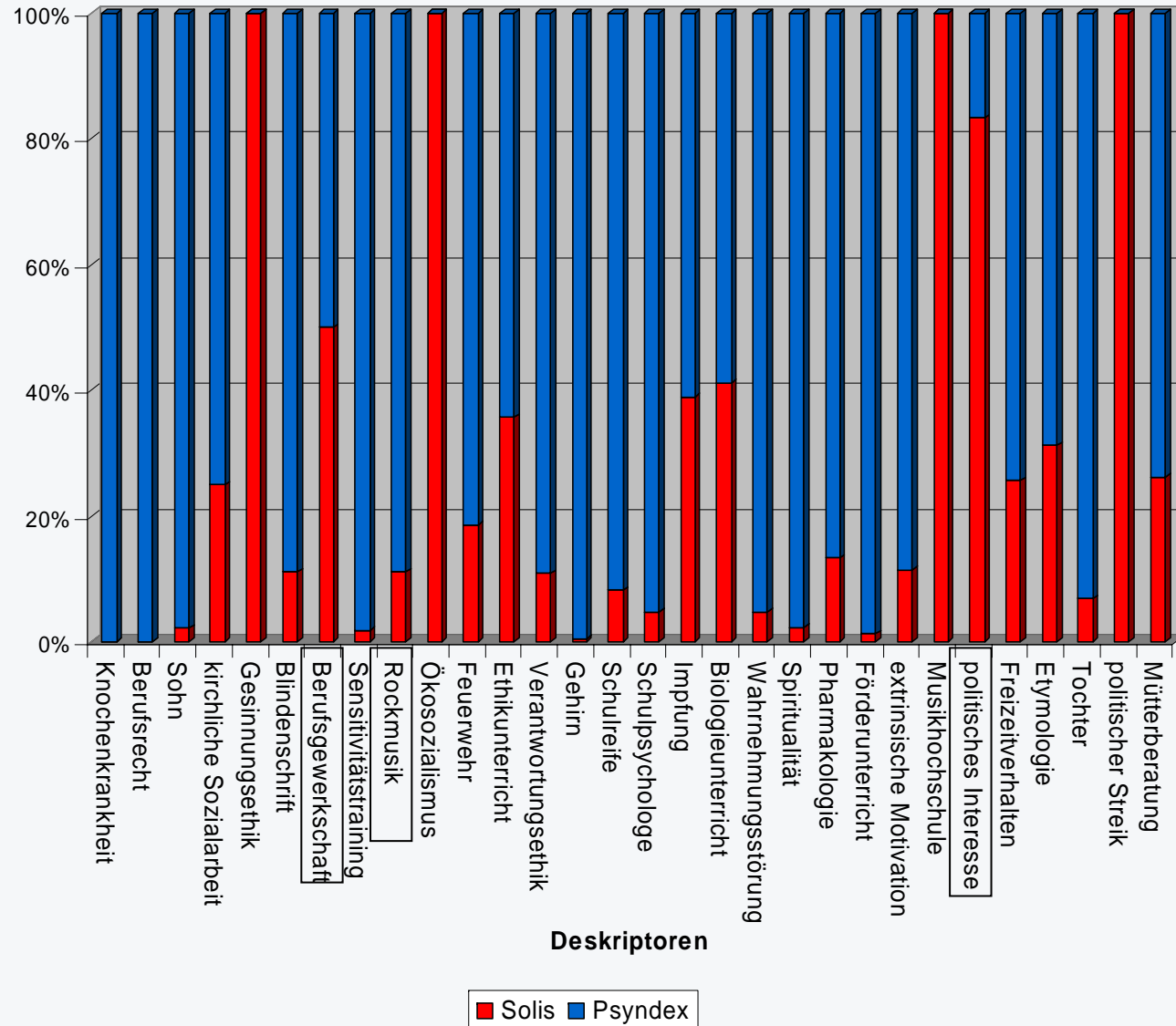


## 3.4.02 Prävention, Therapie, medizinische Rehabilitation

## 30 TheSoz-Deskriptoren, mit denen am häufigsten indexiert wird



## 30 TheSoz-Deskriptoren, mit denen am seltensten indexiert wird





## Zusammenfassung zum Exkurs

- Treffererhöhung kann als nachgewiesen gelten, besonders in interdisziplinären Bereichen
- Vereinfachter Zugang (1 Vokabular)
- noch keine Aussagen zur Relevanz der zusätzlichen Treffer



## Crosskonkordanz-Aufträge

Ausgangsthesaurus	Ausgangsmenge	Zielthesaurus	Zielmenge
Deskriptorenliste BISp	7400	Thesaurus Bildung	55000
Deskriptorenliste BISp	7400	Psyndex Terms	6000
Deskriptorenliste BISp	7400	Thes. Sozialwissenschaften	7500
Psyndex Terms	6000	Deskriptorenliste BISp	7400
Standard Thes. Wirtschaft	5700	Thes. Sozialwissenschaften	7500
Thes. Sozialwissenschaften	7500	Standard Thes. Wirtschaft	5700

Übersicht über die aktuell bearbeiteten Crosskonkordanzen im Teilprojekt:

- Sport → Bildung
- Sport → Psychologie
- Sport → Sozialwissenschaften
- Psychologie → Sport
- Wirtschaft → Sozialwissenschaften
- Sozialwissenschaften → Wirtschaft



## Ausblick

- 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 ( $\wedge$ ,  $>$ ,  $<$ )
  - Z.B. Gestaltung von Benutzeroberflächen
- Recall- und Precision-Untersuchungen (z.B. mit GIRT-Daten)



## Kontakt

Vielen Dank für die Aufmerksamkeit!

Philipp Mayr, Maximilian Stempfhuber & Anne-Kathrin Walter

Informationszentrum Sozialwissenschaften (IZ)  
Abt. Forschung und Entwicklung

Lennéstr. 30  
53113 Bonn  
Tel. 0228 / 22 81 - 0

email {mayr,stempfhuber,walter}@bonn.iz-soz.de  
<http://www.gesis.org/IZ>