

Mapping Knowledge Organization Systems – ein Praxisbericht

*Philipp Mayr, Anne-Kathrin Walter
Informationszentrum Sozialwissenschaften, Bonn*

10. Tagung der Deutschen Sektion der ISKO, 2006



Agenda

- Einführung
 - KOS
 - Semantische Heterogenität
- Projekt KoMoHe
- Mapping von KOS: Crosskonkordanzen
 - Übersicht
 - Praxisbericht
 - Einsatz-Szenario
- Ausblick

Definition:

Knowledge Organization Systems (KOS) is a general term referring to the tools that present the organized interpretation of knowledge structures. Three general categories: Term lists, classifications, relational vocabularies.

(Zeng&Chan, 2004)

“... coined by the Networked Knowledge Organization Systems Working Group at its initial meeting at the ACM Digital Libraries 98 Conference” (Hodge, 2000)

28 Compatibility and Concordances between Indexing Languages

281 Objectives and Nature of Systems Compatibility (10)

282 Intermediate Languages (1)

283 Compatibility in Classing and Indexing (3)

284 Establishment of Concordances (5)

285 Correlative Indexes. Mapping

286 Systems Reconciliation, e.g. between Classification Systems and Thesauri, Linking Terms (1)

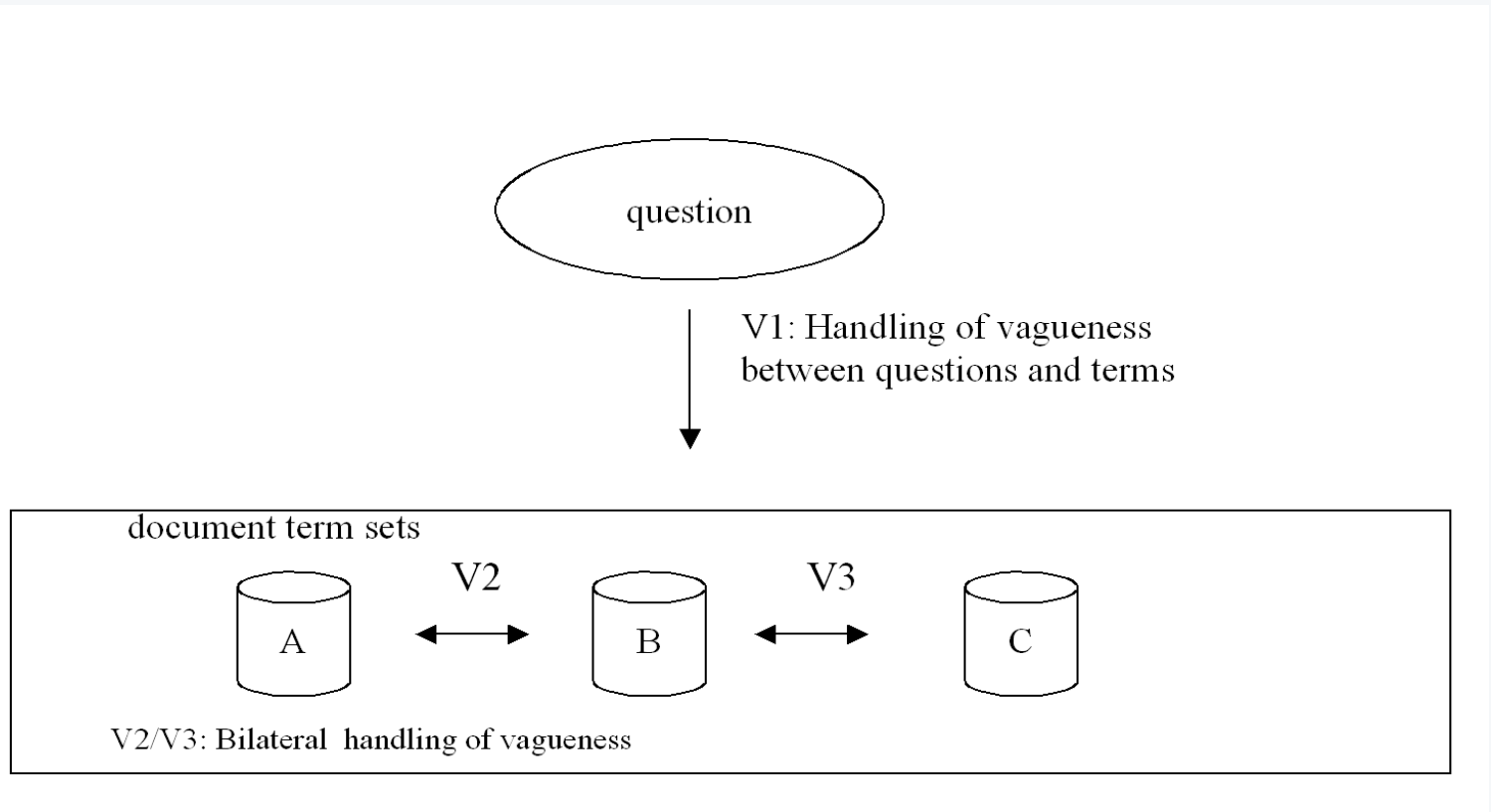
287 Organised Compilation of Compatible Classification Systems and Thesauri, Integration

288 Compatibility in Subject Areas* (1)

289 Evaluation of Compatibility

(aus Classification System for Knowledge Organization Literature, KO 26(1999) No. 4)

Bilateraler Transfer auf Basis der Indexierungssprachen



aus Hellweg et al., 2001

Heterogenitätsbehandlung – Beispiel: infoconnex

- Metasuche 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)

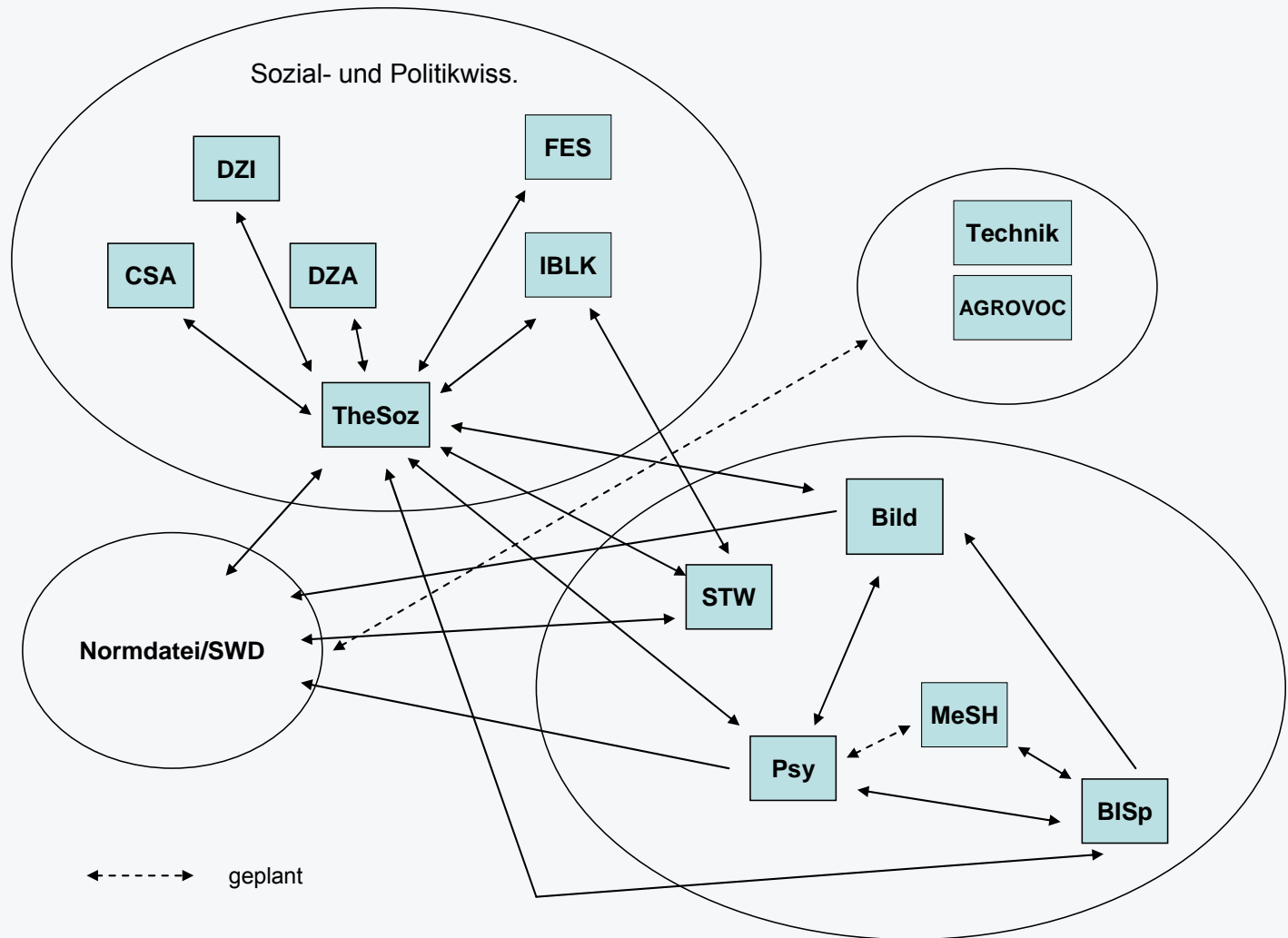
Aktuell nur Verwendung der Äquivalenzrelationen.

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

Crosskonkordanzen Übersicht



1. Vorgehensweise bei der Crosskonkordanz-Erstellung
2. Praxisfälle
3. Einsatzszenario

Ziel: **Äquivalenzrelation** zwischen Ausgangs- und Zielterm

STW	Relation	Relevanz	TheSoz
Nebentätigkeit	=	h	Nebenbeschäftigung
Steuer	=	h	Steuern

Falls nicht möglich, Abarbeitung folgender Schritte:
Kombination aus zwei Termen

STW	Relation	Relevanz	TheSoz
Parteienfinanzierung	=+	h	Partei + Finanzierung
Presseprodukt	=+	h	Presse + Produkt

Suchen nach einem Oberbegriff

STW	Relation	Relevanz	TheSoz
Vogel	<	m	Tier
Betriebssystem	<	m	Software

Suchen nach Unterbegriffen

STW	Relation	Relevanz	TheSoz
Körperpflegemittel	>	m	Kosmetik
Lehrmaterial	>	m	Lehrbuch

Suche nach ähnlichen Termen.

STW	Relation	Relevanz	TheSoz
Geburtenrate	^	h	Geburtenentwicklung
Einlage	^	m	Eigenkapital

Keinen passenden Term gefunden: Nullrelation

STW	Relation	Relevanz	TheSoz
Kupfer	0		
Wirtschaftszone	0		

Vorgehensweise Fortsetzung

- Verweisungen in KOS beachten
- Scope Notes berücksichtigen
- Scope der DB wird nicht in CK abgebildet

Beispiel: Scope DZA-Datenbank: “Gerontologie”

	DZA	Relation	Relevanz	TheSoz
	Tagespflege	=	h	Tagespflege
NICHT:	Tagespflege	=+	h	Tagespflege + alter Mensch

- Mappings nicht zu weit assoziieren
 - Informationsbedürfnis des Nutzers ist i.d.R. vage
 - Abwägung: Recall vs. Precision der Termtransformation

- Einheitliche Verknüpfung von Wortgruppen

Beispiel:

PsyT	Relation	Relevanz	Sport
Tötungsdelikt	<	m	Kriminalität
Gewaltverbrechen	<	m	Kriminalität
Behindertendiskriminierung	=+	h	Behinderter + Diskriminierung
Berufliche Diskriminierung	=+	h	Beruf + Diskriminierung

- Einheitlicher „Abstand“ bei Oberbegriffen

Beispiel: Sportarten

BISp	Relation	Relevanz	TheSoz
Skilanglauf	<	g	Wintersport
Skilauf, alpiner	<	g	Wintersport
Eisklettern	0		
Eisstockschiessen	0		

Kombinationen von Zieltermen (1:n-Relationen)

- nicht mehr als 2-3 Zielterme kombinieren
- Problem:
 - 0 Treffer, wenn Kombination zu speziell
- Beispiel:

Euro-Thes. (IBLK)	Relation	Relevanz	TheSoz
Wohnungsbaugenossenschaft	=+	m	Genossenschaft + Wohnungsbau
Wohnungsbaugenossenschaft	<0	m	Genossenschaft

Beispiel:

STW	Relation	Relevanz	TheSoz
Auskunftspflicht	\wedge	m	Datenschutz

Schwierigkeiten: Grenzbereiche

1) wann eine Nullrelation verwenden?

STW	Relation	Relevanz	TheSoz
Nobelpreis	\wedge	g	Preisverleihung

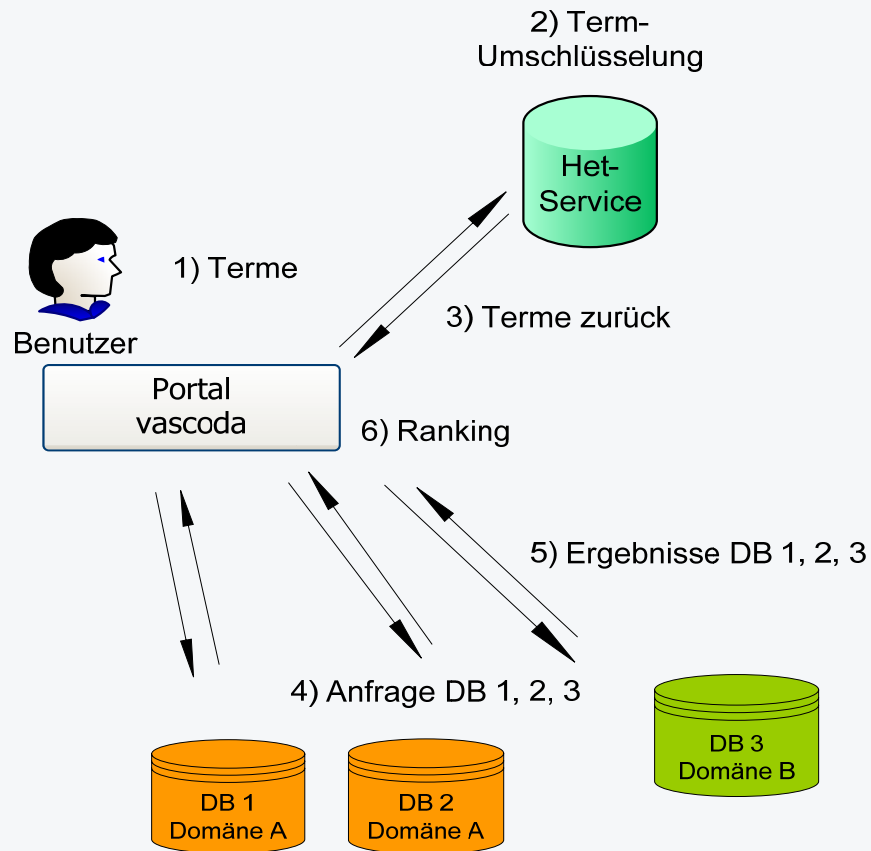
2) wann doch besser eine Äquivalenzrelation benutzen?

STW	Relation	Relevanz	TheSoz
Arbeitsvermittlung	=	h	Arbeitsberatung

SWD	Relation	Relevanz	TheSoz
Anwerbung	\wedge	h	Rekrutierung

Relationen insgesamt	6659
Verknüpfte Terme im Ausgangsthesaurus	5732
Verknüpfte Terme im Zielthesaurus	2968
Äquivalenzrel. (=)	4107
Oberbegriffsrel. (<)	1508
Unterbegriffsrel. (>)	116
Ähnlichkeitsrel. (^)	423
Nullrelationen	505

Einsatzszenario der Crosskonkordanzen



- 1) Nutzer gibt in vascoda Suchterme ein
- 2) Het.-Service schlüsselt übergeb. Terme um
- 3) DB liefert umgeschlüsselte Terme zurück
- 4) Portal setzt Anfrage um und schickt an DB
- 5) Ergebnislieferung pro Datenbank an Portal
- 6) Aufbau der Trefferliste im Portal

- Weitere Crosskonkordanzen
 - Technik
 - Agrovoc
 - Klassifikationen (?)
- Einsatz statistischer Verfahren
 - MeSH-SWD
- Einsatz des Heterogenitätsservice in vascoda, Fachportale
- Anfrageabarbeitung an Benutzerschnittstelle
- Qualitative Evaluierung der Termtransformationen



Vielen Dank für Ihre Aufmerksamkeit!

Philipp Mayr, Anne-Kathrin Walter

{walter,mayr}@bonn.iz-soz.de

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

Crosskonkordanzen:

Gerichtete, relevanzbewertete Relationen zwischen Termen zweier Thesauri oder Klassifikationen

- Erstellung mit Hilfe von Excel-Tabellen
- Relationen:
 - Äquivalenzrelation
 - Oberbegriffsrelation
 - Unterbegriffsrelation
 - Ähnlichkeitsrelation
- Relevanz (bezogen auf die Treffermenge):
 - Hoch,
 - Mittel,
 - Gering
- 1:1- oder 1:n-Verknüpfungen

- Grundlage: Parallelkorpus
 - Bestand, dessen Dokumente mit zwei Indexierungssprachen erschlossen sind
- Messen der Term-Kookkurrenzen:
 - Generierung von Termpaaren pro Dokument
 - Messen der Häufigkeit gleicher Termpaare im Korpus
 - Ermitteln von Gewichtungen für Termpaare (keine Relationen)
- oberer und unterer Schwellenwert



Heterogenitätsservice – Anwendungsszenarien

- einfache Termtransformationen
- in Kombination mit Thesaurus:
Rechercheunterstützung
- Indirekte Transformationen:
 - Transformation über mehrere Schritte
 - Kombination von Transformationsverfahren (Statistische Verfahren und Crosskonkordanzen)
- Transformation einer Kombination von Anfragetermen

Regel für Äquivalenzrelationen

Zeichengleicher Nichtdeskriptor impliziert Äquivalenzrelation.

Beispiel:

STW	Relation	Relevanz	TheSoz
Geistlicher	=	h	(Geistlicher USE) Pfarrer
Geldumlauf	=	h	(Geldumlauf USE) Geldmenge

Indirekte Termtransformationen

