

Zur Methodik von Crosskondordanzen zwischen terminologischen Ressourcen

Berliner Herbsttreffen zur Museumsdokumentation,
DMB Fachgruppentagung,
07. Oktober 2008

Gliederung

1. Einführung
2. Methodik zur Erstellung von Crosskonkordanzen
3. Vernetzung mehrerer terminologischer Ressourcen
4. Ausblick

I. Einführung

Begriffsdefinition

Crosskonkordanzen sind Verbindungen zwischen terminologischen Ressourcen.

Bei der Verknüpfung zweier Vokabulare, auch Mapping genannt, werden die Begriffe zweier terminologischer Ressourcen zueinander in Beziehung gesetzt.

I. Einführung

Relationstypen

„Gebräuchliche“ Relationstypen zw. Begriffen

- Äquivalenzrelation
- Oberbegriffsrelation
- Unterbegriffsrelation

- Assoziationsrelation

Relationen des SKOS Mapping Vocabulary

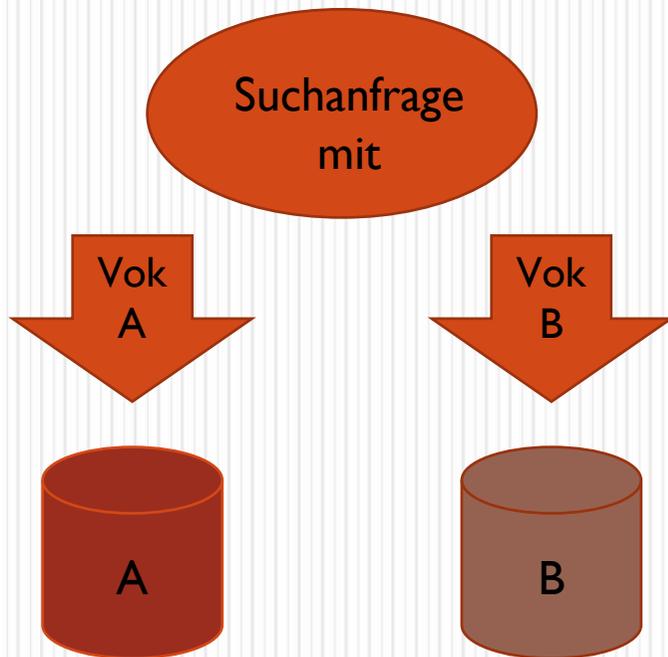
- “has exact match“
- “has broad match“
- “has narrow match“

- “has major match“
- “has minor match“

I. Einführung

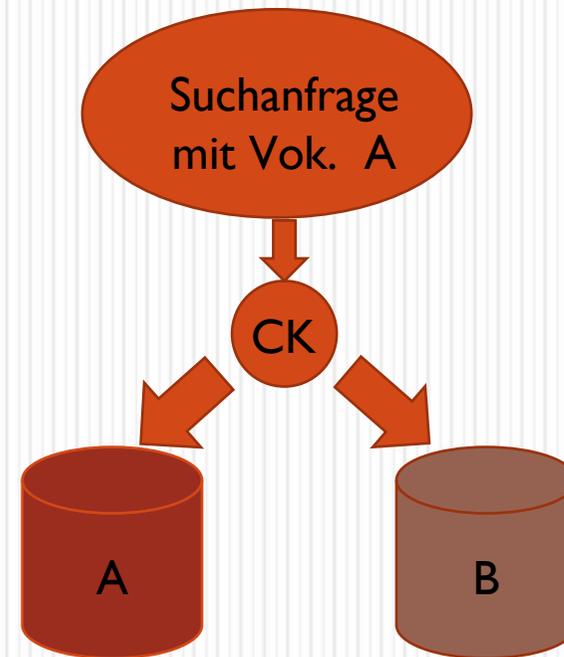
Einsatz von Crosskonkordanzen

- Bisher:



- Zweifache Suchanfrage bei der Suche in zwei Beständen mit unterschiedlichem Vokabular

- Mit Crosskonkordanz:



- Einmalige Suchanfrage - die CK „übersetzt“ die Anfrage mit Vok. A für die Suche in Bestand B

2. Methodik

Allgemeine Aspekte

- Vergleich der Begriffe und Zuordnung der jeweiligen Relationstypen
- Crosskonkordanzen sollten bidirektional sein → beide Vokabulare können als Anfragevokabular genutzt werden
- Crosskonkordanzen sollten bilateral erstellt werden → die Verknüpfung erfolgt vom Ausgangs- zum Zielvokabular und anschließend in umgekehrter Richtung

2. Methodik

Allgemeine Aspekte

- Berücksichtigung der Anwendung der Vokabulare als Indexierungsvokabulare wird empfohlen → Beim Mapping Einsicht in Datensätze, wo diese zur Anwendung kommen
- !Wichtig: Planung der zukünftigen Pflege und Aktualisierung der Crosskonkordanz

2. Methodik

Verfahren zur Crosskordanzerstellung

1. Rein intellektuelles Verfahren
2. Programmgestütztes Verfahren
3. Statistisches Verfahren
4. Ansätze für weitere Verfahren

2. Methodik

Rein intellektuelles Verfahren

=Manuelle Verknüpfung

- Die Begriffssysteme werden analysiert, miteinander verglichen und anschließend in Beziehung gesetzt
 - Gegebenenfalls erfolgt das Einsehen von Datensätzen, in denen die Vokabulare als Indexierungsvokabulare benutzt wurden
- Sehr zeit- und kostenintensiv!

2. Methodik

Programmgestütztes Verfahren

- Einsatz von Programmen zum Zeichenkettenabgleich
 - Entwicklung spezifischer Programme (z.B. Crosskonkordanz Wirtschaft)
 - Anschließend intellektuelle Nachbearbeitung
- Effektiveres Verfahren!

2. Methodik

Statistisches Verfahren

- Grundvoraussetzung: doppelt indexierte Dokumentbestände als Trainingsdaten
 - Mittels Algorithmen Bestimmung der Häufigkeiten des Zusammenvorkommens von Begriffen
 - Aus errechneten Werten zwischen 0 und 1 kann der Grad der Verwandtschaft ermittelt werden
 - Abschließende intellektuelle Nachbearbeitung
-
- Effektives Verfahren
 - Vielversprechender Ansatz
 - Anwendbarkeit für Museumsvokabulare testen

2. Methodik

Ansätze für weitere Verfahren

- Untersuchungen zum Einsatz deduktiver Verfahren
- Erste Versuche zum Einsatz neuronaler Netzwerke

3. Vernetzung

Verbindung mehrerer Vokabulare

1. Direkter Transfer
2. Indirekter Transfer
3. Transfer mittels Zwischenvokabular

3. Vernetzung

Direkter Transfer

- Zwischen allen Vokabularen werden Crosskonkordanzen erstellt
- Bsp.: Vier Thesauri A, B, C und D

Thesaurus A ↔ Thesaurus B Thesaurus B ↔ Thesaurus C
Thesaurus A ↔ Thesaurus C Thesaurus B ↔ Thesaurus D
Thesaurus A ↔ Thesaurus D Thesaurus C ↔ Thesaurus D

- Idealfall für Verknüpfung mehrerer Vokabulare, jedoch sehr aufwändig und kostenintensiv

3. Vernetzung

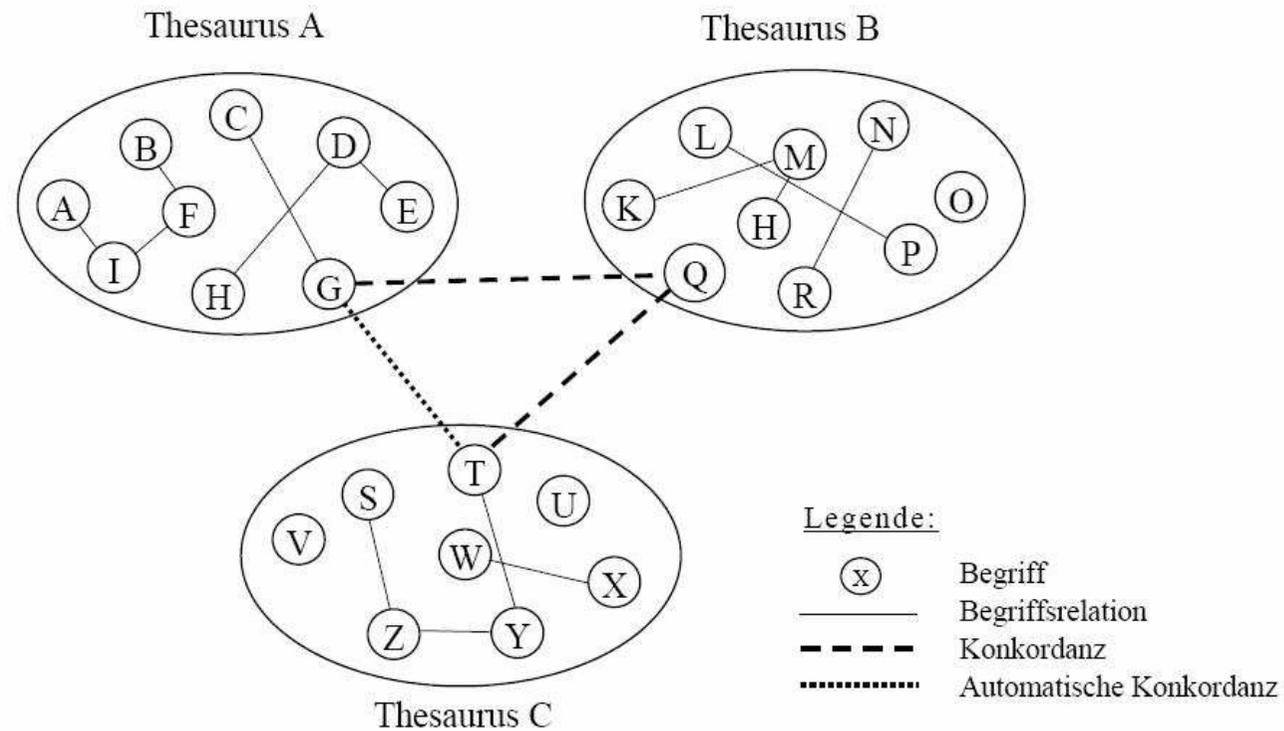
Indirekter Transfer

- Die Crosskonkordanzen werden zwischen „benachbarten“ terminologischen Ressourcen erstellt
- Bsp.: Drei Thesauri A, B und C

Thesaurus A ↔ Thesaurus B ↔ Thesaurus C

3. Vernetzung

Indirekter Transfer



Eigene Darstellung in Anlehnung an: Bernd HERMES 2000, S.25.

3. Vernetzung

Transfer mittels Zwischenvokabular

- Die Begriffe der jeweiligen terminologischen Ressourcen werden mit den Begriffen eines Zwischenvokabulars verknüpft
- Bsp.: Vier Thesauri A, B,C und D

Thesaurus A ↔ Zwischenvokabular X

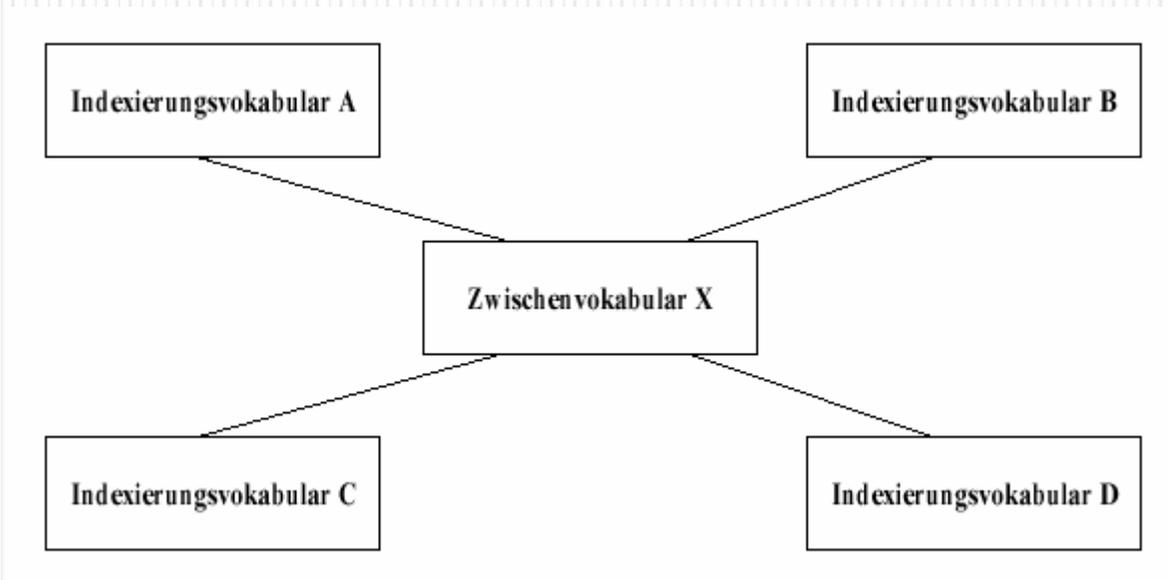
Thesaurus B ↔ Zwischenvokabular X

Thesaurus C ↔ Zwischenvokabular X

Thesaurus D ↔ Zwischenvokabular X

3. Vernetzung

Transfer mittels Zwischenvokabular



Darstellung entnommen aus: Bernd HERMES 2000, S.26.

4. Ausblick

Vorteile und Nebeneffekte

- Vereinfachung und Beschleunigung der mehrere Dokumentbestände übergreifenden Suchanfrage
- Der Fachwissenschaftler kann bei der Suche das ihm vertraute Vokabular benutzen
- Die Vokabulare bleiben unabhängig; deren Pflege bleibt weiterhin in den Händen der Experten

4. Ausblick

Vorteile und Nebeneffekte

- Aufdecken von fehlenden oder falsch zugordneten Begriffen in den terminologischen Ressourcen
- Übereinstimmungen zwischen miteinander konkurrierenden Vokabularen werden ersichtlich

4. Ausblick

Anlaufstellen für weitere Informationen



GESIS IZ Bonn www.social-science-geis.de/iz/
Jürgen Krause, Philipp Mayr



Deutsche Nationalbibliothek www.d-nb.de
Yvonne Jahns

Herzlichen Dank für Ihre
Aufmerksamkeit !

Maria Klingner