UNIVERSITÄT BIELEFELD
Universitätsbibliothek

4. ORCID DE Workshop –
„Organization Identifiers“

Organization Identifier in wissenschaftlichen Einrichtungen

Institutionen-Kodierung für das Kompetenzzentrum Bibliometrie

PD Dr. Niels Taubert
Christopher Lenke

02. Dezember 2020

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Struktur

1. Kompetenzzentrum Bibliometrie (KB)
2. Institutionen-Kodierung als Beitrag zum KB
3. Textmuster
4. Zuordungsquoten
5. Tabellenschema
6. Künftige Entwicklungen

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

kompetenzzentrum bibliometrie
competence centre for bibliometrics

Kompetenzzentrum Bibliometrie (KB)

Ziele

- Bereitstellung einer qualitätsgesicherten Dateninfrastruktur
- Datenbanken Scopus (Elsevier) und Web of Science (Clarivate Analytics)
- Nutzung zur (Weiter-) Entwicklung von Analysemethoden und Indikatoren
- Analyse und Monitoring des deutschen Wissenschaftssystems vermittels bibliometrischer Indkatoren
- Bereitstellung von Daten für die Wissenschafts- und Hochschulforschung (z.B. im Rahmen der BMBF-Förderlinie „Quantitative Wissenschaftsforschung“)

Bundesministerium für Bildung und Forschung

Gefördert vom BMBF Förderkennzeichen: 01PQ17001.

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

kompetenzzentrum bibliometrie
competence centre for bibliometrics

Kompetenzzentrum Bibliometrie (KB)

BMBF geförderter institutionenübergreifender Verbund unter Beteiligung folgender Partner:

- Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW) (Geschäftsstelle),
- Fraunhofer-Institut für System- und Innovationsforschung (Fh-ISI),
- Leibniz-Institut für Informationsinfrastruktur FIZ Karlsruhe (FIZ) (Hosting),
- Forschungszentrum Jülich GmbH (FZJ),
- Max-Planck-Gesellschaft (Max-Planck-Digital Library (MPDL),
- GESIS – Leibniz-Institut für Sozialwissenschaften,
- AG Bibliometrie, Universität Bielefeld (Institutionen-Kodierung)

DZHW Deutsches Zentrum für Hochschul- und Wissenschaftsforschung

Fraunhofer ISI

Universität Bielefeld
Institute for Interdisciplinary Studies of Science (I²SoS)

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

JÜLICH FORSCHUNGSZENTRUM

Max Planck digital library

gesis Leibniz-Institut für Sozialwissenschaften

Bundesministerium für Bildung und Forschung

Gefördert vom BMBF Förderkennzeichen: 01PQ17001.

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Institutionen-Kodierung als Beitrag zum KB

Datenbanken

- Prozessierte Rohdatenbanken (fortlaufend aktualisiert)
- Bibliometriedatenbanken (Stand KW 17)

Versionen der Institutionen-Kodierung

- Jährlich jeweils zwei Kodierung für die Rohdatenbanken und Bibliometriedatenbank (WoS und Scopus)
- Die aktualisierten Versionen der jeweiligen Kodierung erscheinen halbjährlich

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Institutionen-Kodierung als Beitrag zum KB

- **Institutionen-Kodierung**: Systematische Erfassung der institutionellen Einheiten mit ihren Hierarchiebeziehungen unter Berücksichtigung der Veränderungen in der Zeit (Historisierung)
- **Adress-Kodierung**: Vorgang der Zuordnung von konkreten Autor*innen-Adressen zum einschlägigen Institutionen-Schlüsseln
- **Zusätzlich**: Daten zu den Institutionen und Sektoren der deutschen Forschungslandschaft, soweit relevant für die in den Datenbanken verzeichneten Publikationen haben.
- **Realisierung**: Semi-automatisches Verfahren, das wesentlich auf der Erkennung von Textmustern in Adressen beruht.

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Institutionen-Kodierung als Beitrag zum KB

- Adress-Daten mit Publikationsjahr und Country-Code

Konzentration auf einen Teil der Adressen

Transformationsregeln

Laden deutscher Adressen

Erstellung des Adress-Strings

Transformation

```
address='Universität Bielefeld, Fakultät für Mathematik, D-33615 Bielefeld, DEU'
city='Bielefeld'
Pubyear=2010
```

```
address='UNIV BIELEFLED, FAK MATH, D-33615 BIELEFELD, DEU '
city='BIELEFELD'
```

Vorbereitende Datentransformation

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Institutionen-Kodierung als Beitrag zum KB

- Regular Expressions als Zuordnungskriterien

PublikAdress-Daten mit

Regular Expressions als Ausschlusskriterien

Code

ationsjahr und

Sektorzuordnung (hierarchische Beziehung)

Hierarchische Beziehung innerhalb von Organisationen

Anwendung Textmuster

Anwendung Ausschlussmuster

Zuordnung zur Residualkategorie

Aggregation auf niedrigsten Zuordnungslevel

```
address='UNIV BIELEFLED, FAK MATH, D-33615 BIELEFELD, DEU '
city='BIELEFELD'
```

```
Adress-Muster | City-Muster | ID
UNIV BIELEFELD% | % | 170
%UNIV BIELEFELD% | BIEL%FELD | 170
%FAK MATH% |%BIEL%FELD% | 1809
```

```
Ausschlussmuster greifen nicht, gefundene Zuordnung wird beibehalten
```

```
Es findet keine Zuordnung zur Residualkategorie statt, da konkrete Zuordnung vorgenommen wurde
```

```
Zuordnung zu Kategorie 1809, da es sich um das niedrigste Zuordnungslevel handelt
```

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Textmuster

Beispiel: RWTH Aachen

-

Adressmuster	Citymuster
AACHEN UNIV%	%AACHEN%
%UNIV TECH AIX LA CHAPELLE%	%
THAACHEN%	%
RHEIN WESTFAL TECH HSCH%	%AACHEN%
WESTFAL TECH HSCH,%	%AACHEN%
RTW AACHEN UNIV,%	%AACHEN%
AACHEN TECH HSCH,%	%AACHEN%
RHINE WESTFALIA TECH UNIV%	%AACHEN%
TECH HOCHSCHULE AACHEN,%	%AACHEN%
RWTH AACHEN%	%AACHEN UNIV%

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Textmuster

Kriterien zur Erstellung von Textmustern

- So einfach wie möglich (Vermeidung von Fehlern, Überblick)
- So allgemein wie möglich (Ziel: hoher Recall)
- So spezifisch wie nötig (Ziel: hohe Precision)

Textmuster und Institutionen

- Anzahl Textmuster > 52.000
- Anzahl erfasster Institutionen in Deutschland: 2.151
- Davon existierende Institutionen: 2.077

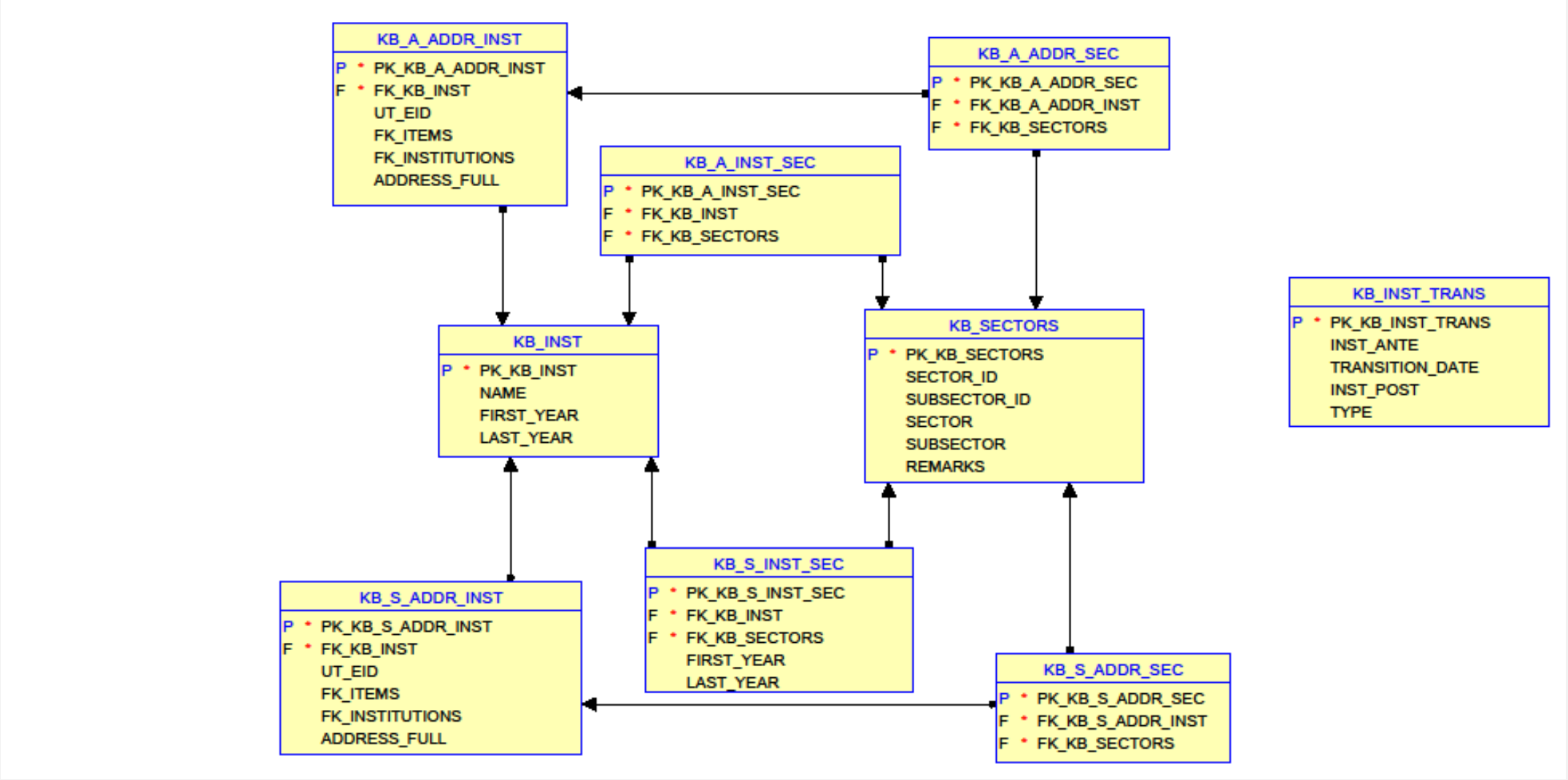
UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Zuordungsquoten

	WoS BDB Kodierung 08/2020	Scopus BDB Kodierung 10/2020
Anzahl distinkte Adressen	2.377.695	3.254.876
Anzahl distinkte Adressen, zugeordnet	2.048.334 (86,15%)	2.733.361 (83,98%)
Anzahl distinkte Adress-Dokument-Kombinationen	7.168.927	5.800.126
Anzahl distinkte Adress-Dokument-Kombinationen, zugeordnet	6.672.948 (93,08%)	5.093.284 (87,81%)
Anzahl distinkte Dokumente	3.884.483	3.325.562
Anzahl distinkte Dokumente mit mindestens einer Zuordnung	3.679.229 (94,72%)	3.019.281 (90.79%)

UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Datenbankschema



UNIVERSITÄT BIELEFELD
Universitätsbibliothek

Künftige Entwicklungen

- Anreicherung der Institutionen-Kodierung mit GRID_ID
- Abbildung der amtlichen Statistik in der Institutionen-Kodierung
- Test der Institutionen-Kodierung auf weiteren Datenquellen (z.B. Dimensions, Crossref)
- Entwicklung von Szenarien für eine Nachnutzbarkeit