

Arbeitsbericht AG Indexierung der Konferenz für Regelwerksfragen

Reiner *Diedrichs*, Göttingen

Vor dem Hintergrund der zunehmenden Verbreitung von Suchmaschinen für bibliothekarische Datenbanken – KvK, KOBV, etc. – ergeben sich aus der unterschiedlichen Konstruktion der angegangenen Indexsysteme einige Probleme:

- inkonsistente Suchergebnisse bei scheinbar gleichen Suchschlüsseln
Autor: Müller-Udenscheid -> müller, Udenscheid, mueller?
Titelstichworte: sind Zusätze zum Sachtitel indexiert oder nicht?
- unterschiedliche Suchschlüssel: Beschränkung auf die kleinste gemeinsame Schnittmenge

Die Verwendung einer einheitlichen Oberfläche gaukelt an dieser Stelle aber konsistente Ergebnisse vor.

Auf Grund dieser Ausgangslage hat die KfR eine Arbeitsgruppe eingesetzt, die Empfehlungen für eine einheitliche Indexierung bibliothekarischer Datenbanken erarbeiten soll. Diese Arbeitsgruppe hat sich bisher dreimal getroffen und Entwürfe für erste Festlegungen erarbeitet.

Die vorliegenden Ergebnisse sollen in ihren Grundzügen kurz vorgestellt werden. Nach Verabschiedung durch die KfR bzw. ihrer Nachfolgeorganisation werden diese veröffentlicht.

Arbeitsprogramm

Da der Komplex Indexierung sehr umfangreich und vielschichtig ist, ist die Arbeitsgruppe der Ansicht, dass eine schrittweise Vorgehensweise notwendig ist:

1. Einschränkung auf Titeldaten, auf die RAK Anwendung findet
2. Normdateien, wie PND, SWD
3. Notationssysteme (siehe auch AG Index)

Im ersten Schritt ist festzulegen, welcher Inhalt wie indexiert wird. Berücksichtigt werden die Indextypen Wort, String und Person. Weitere Indextypen bleiben zunächst unberücksichtigt, da sie sehr kontextabhängig sind. Auch die Verwendung von Wörterbüchern etc. können in einer ersten allgemeinen Empfehlung nicht berücksichtigt werden. Ergänzend sind die zugehörigen Bildungsregeln für die o.g. Indextypen und ergänzende Bestimmungen, wie Zeichenumsetzung, Sortierwerte von Zeichen usw., detailliert zu beschreiben.

Begriffsfestlegungen

Um einen einheitlichen Sprachgebrauch zu gewährleisten werden folgende Festlegungen für die Verwendung des Begriffes **Index** an Stelle von Register, invertierte Liste getroffen:

- Index Inhalt: Autoren, Titelstichworte, Schlagwort, etc.
- Indexeintrag
- Indextyp inhaltlich = Name, Wort, etc.: Titelstichworte = Indextyp "Wort"
- Indexroutine technisch formale Bildung = Stichwort, String, Name

Für die Bezeichnung Primär/Sekundär-Aspekt sollen einheitlich benutzt werden:

- Primärer Zugriff: Zugriff über Indexeinträge
- Sekundärer Zugriff: Einschränkung nicht direkt suchbare Aspekte, wie z.B. Jahrgang

Zu indexierende Inhalte

Zielgruppe ist der Endnutzer. Der Umfang der Indexierung ist auf die Bedürfnisse der Nutzer und die darüber hinausgehenden Bedürfnisse der Bibliothekare auszurichten.

- **Basis-Index (All-Index): Indextyp Wort**
 - Nummern (aus allen Nummernfeldern)
 - Titel und Titeltitelzusätze (aus allen Titelfeldern)
 - Personennamen (aus allen Namensfeldern)
 - Körperschaften
 - Kongresse
 - Gesamttitel
 - Sacherschließung (Schlagworte)
 - Abstracts
- **Personen: Indextyp Person**
alle = Person als Verfasser etc. (von der Person), auch als Schlagwörter (über die Person).
- **Titel: Indextyp Wort**
- alle Titelfelder, Titel der Gesamtaufnahme, (nicht Gesamttitel Feld 451); Zusätze zum Sachtitel; zu ergänzender Urheber, sachliche Benennung von Bänden mehrbändiger Werke
- **Titel: Indextyp String**
Titel und zu ergänzender Urheber (Titeltitelzusätze sind noch zu klären).
- **Gesamttitel: Indextyp String**
Gesamttitel-Hauptaufnahmen mit Zählung
- **Nummern: Indextyp Wort**
Alle Nummern, wie ISBN, Reportnummern, ISSN
- **Körperschaften: Indextyp Wort**
- **Körperschaften: Indextyp String**
- **Verlag: Indextyp Wort**
- **Ort: Indextyp Wort**
- **Schlagwort (noch offen)**

Für den sekundären Zugriff

- **Codes** (Festlegungen erfolgen nach Verabschiedung der Ergebnisse der Arbeitsgruppe)
- **Jahr**
- **Erscheinungsverlauf** (offen)

Der Entwurf für genaue Zuordnung der Felder liegt auf MAB2 bezogen vor.

Bildungsregeln für Indextypen

Person

Ein Arbeitspapier auf Basis der PND-Regeln liegt vor. Im Detail zu klären ist die Behandlung von Präfixen und Verwandtschaftsbezeichnungen im Zusammenhang mit der Bildung von Rotationsindices und Normierung von Ansetzungen. Zu erarbeiten ist hier eine Liste von einheitlichen Wendungen, die eine maschinelle Bearbeitung des Problems erlaubt.

Wort

Ein detaillierter Entwurf der Bildungsregeln liegt vor. Geregelt wird dort u.a., dass

- nach Entfernen der Sonderzeichen alle Buchstaben in Kleinbuchstaben umgesetzt werden,
- Umlaute umgesetzt werden nach “Grundbuchstabe + e” und “Grundbuchstabe”,
- möglichst keine Stopwortlisten verwendet werden,
- griechische Buchstaben durch die lateinische Bezeichnung ersetzt werden,
- Elemente in [...] entfernt werden,
- Nicht-Sortierzeichen entfernt werden.

String

Ein detaillierter Entwurf liegt vor. Wesentliche Regelungen sind:

- Elemente in Nicht-Sortierzeichen werden i.d.R. entfernt.
- Ergänzungen in [...] werden entfernt.
- Buchstaben werden in Kleinbuchstaben umgesetzt.

Weitere Vorgehensweise

Bedingt durch die Arbeitsbelastung einzelner Mitglieder der Arbeitsgruppe ist die Zusammenfassung der Themenkomplexe zu einem Gesamtentwurf noch nicht fertig, wird aber in nächster Zeit realisiert. Dieser Entwurf wird voraussichtlich in einer abschließenden Sitzung fertiggestellt und dann der KfR vorgelegt. Die Veröffentlichung soll noch im Jahr 2000 erfolgen.

