

# **Methoden der Inhaltsererschließung I. Indexierung**

## **Grundlagen**

Elena Semenova

FH Potsdam

# Worum handelt es sich hierbei?

Rösener, Christoph: Die Stecknadel im Heuhaufen.  
Peter Lang Europäischer Verlag der Wissenschaften,  
Frankfurt am Main, Berlin, Bern, Bruxelles, New York,  
Oxford, Wien, 2005, ISBN 978-3-631-54447-1

Ein Märchen?

Eine Erzählung?



Ein Roman?

Abhandlung über die  
deutsche Sprichwörter?

...???

**Problem**

**Lösung: Indexierung**

<b>Link zu diesem Datensatz</b>	<a href="http://d-nb.info/97628300X">http://d-nb.info/97628300X</a>
<b>Titel</b>	Die Stecknadel im Heuhaufen : natürlichsprachlicher Zugang zu Volltextdatenbanken / Christoph Rösener
<b>Person(en)</b>	Rösener, Christoph
<b>Verleger</b>	Frankfurt am Main ; Berlin ; Bern ; Bruxelles ; New York ; Oxford ; Wien : Lang
<b>Erscheinungsjahr</b>	2005
<b>Umfang/Format</b>	X, 243 S. : Ill., graph. Darst. ; 21 cm
<b>Gesamttitle</b>	Saarbrücker Beiträge zur Sprach- und Translationswissenschaft ; Bd. 8
<b>Hochschulschrift</b>	Zugl.: Saarbrücken, Univ., Diss., 2005
<b>ISBN</b>	3-631-54447-2 
<b>EAN</b>	9783631544471 
<b>Einband/Preis</b>	kart. : EUR 45.50 (DE), EUR 45.50 (AT)
<b>Schlagwörter</b>	Brockhaus-Enzyklopädie ; Volltextdatenbank ; Retrievalsprache ; Natürliche Sprache ; Visualisierung Information Retrieval ; Wissensrepräsentation ; Automatische Sprachanalyse ; Automatische Indexierung ; Dialogsystem
<b>Sachgruppe(n)</b>	004 Informatik ; 020 Bibliotheks- und Informationswissenschaft ; 400 Sprache, Linguistik

<b>Frankfurt</b>	Signatur: 2006 A 77728 Bereitstellung in Frankfurt
<b>Leipzig</b>	Signatur: 2006 A 21401 Bereitstellung in Leipzig

# Indexierung. Definition

- Der Vorgang der inhaltlichen oder sachlichen **Indizierung** von Dokumenten für eine **gezielte Wiedergewinnung** mit frei wählbarem oder gebundenem Vokabular. [Semiotischer Thesaurus, HU Berlin]
- Als **Indexierung**, auch **Verschlagwortung** (Österreich: Beschlagwortung) oder **Verstichwortung**, bezeichnet man beim **Information Retrieval** die Zuordnung von **Deskriptoren** zu einem Dokument **zur Erschließung** der darin enthaltenen Sachverhalte. [Wikipedia]

## Indexierung

UF Verschlagwortung

UF Verstichwortung

UF Beschlagwortung

BT Verfahren der Information Retrieval

BT Verfahren der Inhaltlichen Erschließung

RT Dokumentationsprache

# Zweck der Indexierungsverfahren



**Erfolgreiche  
Recherche**

*Kriterien:*

- **Recall**
- **Precision**

# Indexterme

Semenova. Grundlagen der Erschließung

Link zu diesem Datensatz	<a href="http://d-nb.info/97628300X">http://d-nb.info/97628300X</a>
Titel	Die Stecknadel im Heuhaufen : natürlichsprachlicher Zugang zu Volltextdatenbanken / Christoph Rösener
Person(en)	Rösener, Christoph
Verleger	Frankfurt am Main ; Berlin ; Bern ; Bruxelles ; New York ; Oxford ; Wien : Lang
Erscheinungsjahr	2005
Umfang/Format	X, 243 S. : Ill., graph. Darst. ; 21 cm
Gesamttitle	Saarbrücker Beiträge zur Sprach- und Translationswissenschaft ; Bd. 8
Hochschulschrift	Zugl.: Saarbrücken, Univ., Diss., 2005
ISBN	3-631-54447-2
EAN	9783631544471
Einband/Preis	kart. : EUR 45.50 (DE), EUR 45.50 (AT)
Schlagwörter	Brockhaus-Enzyklopädie ; Volltextdatenbank ; Retrievalsprache ; Natürliche Sprache ; Visualisierung ; Information Retrieval ; Wissensrepräsentation ; Automatische Sprachanalyse ; Automatische Indexierung ; Dialogsystem
Sachgruppe(n)	004 Informatik ; 020 Bibliotheks- und Informationswissenschaft ; 400 Sprache, Linguistik

**Stichwort**

**Schlagwort**

# Indexterme

- **Stichwort**

*ist ein aus einem gegebenen Dokumenttext **extrahiertes** und diesen kennzeichnendes Wort*

- **Schlagwort**

*ist [...] ein Wort, das einem Text zur inhaltlichen Beschreibung **zugeteilt** wird, ohne notwendigerweise im Text selbst vorkommen zu müssen*

Quelle: Bertram, J.: Einführung in die inhaltliche Erschließung.

**Schlagwort: freies Schlagwort, Deskriptor, Notation**





# Indexierungsmethoden

## Extraktions- und Additionsmethode


### Extraktionsmethode

*Indexterme werden  
dem Dokument  
entnommen  
(Stichwörter)*

### Additionsmethode

*Indexterme werden dem  
Dokument zugeteilt  
(Schlagwörter,  
Stichwörter)*

Höhere Recall und  
Precision

<b>Link zu diesem Datensatz</b>	<a href="http://d-nb.info/954257758">http://d-nb.info/954257758</a>
<b>Titel</b>	Das Aschaffenerburger Tafelbild : Studien zur Tafelmalerei des 13. Jahrhunderts ; München, 8. - 10. Mai 1996 / Internationales Kolloquium zur Tafelmalerei des 13. Jahrhunderts. Veranst. vom Bayerischen Landesamt für Denkmalpflege und dem Zentralinstitut für Kunstgeschichte. Erwin Emmerling/Cornelia Ringer (Hrsg.)
<b>Person(en)</b>	Emmerling, Erwin [Hrsg.]
<b>Organisation(en)</b>	Internationales Kolloquium zur Tafelmalerei des 13. Jahrhunderts <1996, München>
<b>Verleger</b>	München : Lipp
<b>Erscheinungsjahr</b>	1997
<b>Umfang/Format</b>	391 S. : zahlr. Ill. ; 30 cm
<b>Gesamttitel</b>	Bayern: Arbeitshefte des Bayerischen Landesamtes für Denkmalpflege ; Bd. 89
<b>ISBN</b>	3-87490-653-1 
<b>Einband/Preis</b>	kart. : DM 78.00
<b>Schlagwörter</b>	Aschaffenerburger Tafelbild Kongress ; München <1996> Aschaffenerburger Tafelbild, Restaurierung Kongress ; München <1996> Deutschland, Malerei : Geschichte 1200-1300 Kongress ; München <1996>
<b>Sachgruppe(n)</b>	46 Bildende Kunst
<b>Extraktion</b>	Signatur: 1998 B 11622 Bereitstellung in Frankfurt
	Leipzig Signatur: 1998 B 11622 Bereitstellung in Leipzig
<b>Addition</b>	

**Extraktion**

**Addition**



# Indexierungsmethoden

## Intellektuelles vs. automatisches Indexieren

Intellektuelles Indexieren

→ intellektuelle  
Inhaltsanalyse (frei  
oder gebunden;  
Additionsmethode  
bzw.  
Extraktionsmethode)

begriffsorientiert

Höhere Recall und Precision

Automatisches Indexieren

→ maschinelle Verfahren  
(hauptsächlich  
Extraktionsmethode)

benennungsorientiert

# Indexierungsmethoden

## Intellektuelles vs. automatisches Indexieren

*Intellektuelles Indexieren arbeitet mit*



**Referent**

*außersprachliche Realität*

*Automatisches Indexieren arbeitet mit*

## Intellektuelle Erschließung vs. Automatische Indexierung

<b>Ziele</b>	<ul style="list-style-type: none"> <li>• Retrievalverbesserung durch geeignete Indexterme</li> <li>• Recallerhöhung auf Stichwortbasis</li> </ul>	<ul style="list-style-type: none"> <li>• einheitliche thematische Suche durch thematische Zusammenführung</li> <li>• inhaltliche Repräsentation</li> </ul>
<b>Aufwand</b>	• gering	• hoch
<b>Kosten</b>	• gering	• hoch
<b>Ergebnisse</b>	<ul style="list-style-type: none"> <li>• Indexterme</li> <li>• vorhersagbar</li> <li>• beliebig reproduzierbar</li> <li>• durch wiederholte Indexierungen leicht zu verbessern</li> </ul>	<ul style="list-style-type: none"> <li>• Schlagwörter, Deskriptoren, Notationen</li> <li>• menschliche Fehlerquote</li> <li>• endgültig</li> <li>• inhaltlich verlässlich</li> </ul>
<b>Qualität</b>	• abhängig von Qualität der Stichwörter im Quelltext	• hoch, bei konsistenter Erschließung
<b>Nutzen</b>	<ul style="list-style-type: none"> <li>• für alle textbasierten Quelldokumente (auch erschlossene!) hoch</li> <li>• insb. für nicht erschlossene und nicht zu erschließende Dokumente hoch</li> </ul>	<ul style="list-style-type: none"> <li>• bezogen auf die erschlossenen Dokumente hoch</li> <li>• in Kollektionen mit niedriger Erschließungsquote gering</li> </ul>

Automatische Indexierung Einführung

16

Semenova. Grundlagen der Erschließung

# Indexierungsmethoden

## Gleichordnendes vs. syntaktisches Indexieren

Thieme, Cristina: Das Tafelbild aus der Kathedrale von Trogir : kunsttechnologische Studien zur Tafelmalerei Dalmatiens. - Wiesbaden : Reichert, 2007 (2007). - 289 S.

### *Indexat I*

Tafelbild - Dom – Malerei – Restaurierung – Trogir -  
Dalmatien

*Semantische Zusammenhänge sind nicht abgebildet*  
= **gleichordnendes Indexieren**

### *Indexat II*

- Trogir: Dom; Tafelbild
- Dalmatien: Malerei
- Malerei; Restaurierung

*Semantische Zusammenhänge sind syntaktisch wiedergegeben =*  
**syntaktisches Indexieren**

# Indexierungsmethoden

## Gleichordnendes vs. syntaktisches Indexieren

Konsequenzen für Recherche

- Gleichordnendes Indexieren erfordert eine Zusammenfügung komplexer Themen in der Suche

z.B. Trogir AND Dom

→ Postkoordination

- Syntaktisches Indexieren liefert schon im Vorfeld zusammengeführte Sachverhalte

z.B. Trogir: Dom

→ Präkoordination

# Präkombination, Präkoordination, Postkoordination

- **Präkombination** – Verfahren, bei denen ein Sachverhalt **in der Dokumentations-sprache** festgelegt wird
  - z.B. Kroatische Kunst  
OB Kunst
- **Präkoordination** – Verfahren, bei denen ein Sachverhalt **während des Indexierungsprozesses** festgelegt wird
  - z.B. Kroatien: Kunst
- **Postkoordination** – Verfahren, bei denen ein Sachverhalt **während der Recherche** bestimmt wird
  - z.B. Kroatien AND Kunst



# Indexierungsmethoden

## Freies vs. gebundenes Indexieren

### Freies Indexieren

*Indexterm ist FREI  
ausgewählt*

**Vorteile:** schnell,  
kostengünstig

**Nachteile:** Probleme der  
natürlichen Sprache sind  
nicht aufgehoben →  
Verlust bei Recherche

### Gebundenes Indexieren

*Indexterm ist einem  
kontrollierten Vokabular  
entnommen*

**(gebundene Schlagwörter,  
Deskriptoren, Notationen)**

**Vorteile:** normierter Wortschatz  
→ hohe Precision und  
Recall

**Nachteile:** zeit- und kosten  
aufwendig

**Höhere Recall und Precision**