



UNIVERSITÄT
LEIPZIG

UNIVERSITÄTS
BIBLIOTHEK *ubl* LEIPZIG

Aus dem Alltag einer Universitätsbibliothek

Beobachtungen und Gedanken zur Anforderung an bibliographische Metadaten

Metadaten zwischen Autopsie und Automatisierung - welche Qualität brauchen wir?

Fortbildungsveranstaltung der dbv-Kommission Erwerbung und Bestandsentwicklung

ULB Düsseldorf, 13. März 2017

Björn Muschall

Metadaten zwischen Autopsie und Automatisierung

WELCHE QUALITÄT BRAUCHEN WIR? UND WOFÜR?

Metadaten für

INFORMATION RETRIEVAL IN LOKALEN UND GLOBALEN INFORMATIONSRÄUMEN

```
"facet.field": "mega_collection",
"indent": "true",
"start": "0",
"rows": "0",
"wt": "json",
"facet": "true",
"_: "1519998819559"
}
},
"response": {
  "numFound": 136047097,
  "start": 0,
  "docs": []
},
"facet_counts": {
  "facet_queries": {},
  "facet_fields": {
    "mega_collection": [
      "Verbunddaten SWB",
      10114818,
      "Lizenzfreie Online-Ressourcen",
      1096027,
      "Persée",
      712766,
      "PDA Print Nielsen",
      508644,
      "SWB SSG UB Heidelberg",
      494843,
      "Genios (Wirtschaftswissenschaften)",
      17620269,
      "Elsevier BV (CrossRef)",
      14753007,
      "Wiley-Blackwell (CrossRef)",
      7582515,
      "Springer Science + Business Media (CrossRef)",
      4337890,
      "Genios (Recht)",
      11760001
    ]
  }
}
```

Heterogene und homogene

INFORMATIONEN RÄUME

HETEROGENER INFORMATIONSRAUM UBL KATALOG

- fine Infrastruktur
- über 150 Quellen
- über 2.200 Kollektionen
- über 75 Mio. Records (Bestand)
 - davon 4,4 Mio SWB
- Metadaten, Abstracts, TOC, Volltext
- Quellformate: MARC21, MODS, LIDO, JATS, Crossref, HTML, ...
- Anzeigeformate (Fullrecord): MARC21, LIDO, AI-Intermediate
- Titeldaten und Exemplardaten
- Tägliche bzw. monatliche Updates

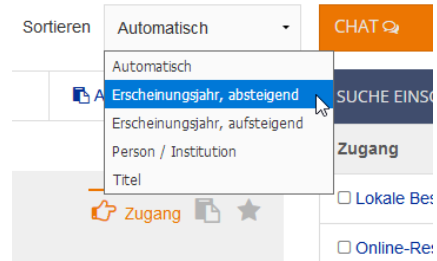
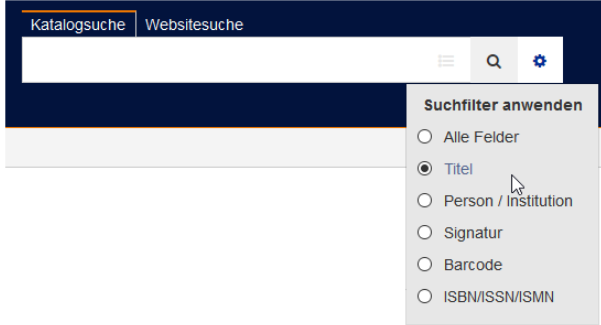
EXTERNE DATENQUELLEN FINC (AUSWAHL)

- Crossref
- Nielsen Book Data, VUB
- DeGruyter, Springer, Wiley,
- Elsevier Journals
- IEEE Xplore Library
- JSTOR
- Juris
- GBI Genios Wiso
- Oxford Scholarship Online
- OECD iLibrary
- Thieme eBooks
- DNB
- Deutsche Fotothek (SLUB)
- Nationallizenzen
- DOAB, DOAJ, SSOAR
- Deutsches Textarchiv
- Naxos Music Library
- RISM
- Qucosa
- ...

HETEROGENITÄT

- Sprache und Schrift
- Ansetzung Autoren
- Verfügbarkeit Normdaten
- Repräsentation des Dokuments: Volltext, Abstract, TOC, Metadaten
- Vollständigkeit / Detailgrad der Metadaten
- Verfügbarkeit Sacherschließung
- Verschiedene (kontrollierte) Vokabulare
- Dokumente aus verschiedenen Wissensdomänen
- Materialart
- Verwendung von Regelwerken

NORMIERTE / NORMALISIERTE SUCHEINSTIEGE IM HETEROGENEN KATALOG



ERWEITERTE SUCHE

Suche nach:

Übereinstimmung:

+ Suchfeld hinzufügen

+ Suchgruppe hinzufügen

Felder leeren

Finden

Begrenze auf

Zugang:

- Lokale Bestände
- Online-Ressourcen

Standorte:

- Allgemeine
- Archäologie
- Bibliotheca
- Biowissens
- Campus-Bi
- Chemie/Ph
- Deutsches Literaturinstitut Leipzig
- Erziehungswissenschaft
- Geographie
- Geowissenschaften

Bereich:

- Freihandbereich
- Lehrbuchsammlung
- Magazin / Sonstige

Treffer pro Seite

20

Erscheinungsjahr

Von:

Bis:

Alle Felder

- Alle Felder
- Titel
- Person / Institution
- RVK-Notation
- Schlagwort
- Serie
- Verlag
- Erscheinungsort
- Erscheinungsjahr
- ISBN / ISSN
- Signatur
- Barcode
- Inhaltsangabe

SUCHE EINSCHRÄNKEN	
Zugang ▲	
<input type="checkbox"/> Lokale Bestände	3.208.510
<input type="checkbox"/> Online-Ressourcen	1.179.231

Format ▲	
<input type="checkbox"/> Buch, E-Book	2.683.930
<input type="checkbox"/> Hochschulschrift	736.715
<input type="checkbox"/> Artikel	524.103
<input type="checkbox"/> Zeitschrift, E-Journal	301.833
<input type="checkbox"/> Noten	51.600
<input type="checkbox"/> Audio	28.515
mehr ...	


Standorte ▼

Fachgebiet ▲	
<input type="checkbox"/> nicht zugeordnet	2.488.864
<input type="checkbox"/> Soziologie	261.487
<input type="checkbox"/> Geschichte	259.614
<input type="checkbox"/> Rechtswissenschaft	170.414
<input type="checkbox"/> Wirtschaftswissens...	163.092
<input type="checkbox"/> Technik	157.703
mehr ...	

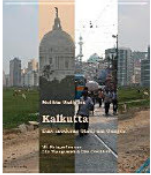
Person / Institution ▲	
<input type="checkbox"/> Deutschland	7.759
<input type="checkbox"/> Association for Comp...	6.044
<input type="checkbox"/> Goethe, Johann Wolfg...	4.419
<input type="checkbox"/> Luther, Martin	4.264
<input type="checkbox"/> Universität Leipzig	3.697
<input type="checkbox"/> Friedrich	3.599
mehr ...	

Sprache ▲	
<input type="checkbox"/> Deutsch	2.689.231
<input type="checkbox"/> Englisch	675.693
<input type="checkbox"/> Französisch	295.307
<input type="checkbox"/> Latein	258.100
<input type="checkbox"/> Russisch	99.192
<input type="checkbox"/> Italienisch	71.736
mehr ...	

Kollektion ▲	
<input type="checkbox"/> Verbunddaten SWB	4.387.742
<input type="checkbox"/> Lizenzfreie Online...	1.096.045
<input type="checkbox"/> SWB SSG UB Heid...	149.179
<input type="checkbox"/> Sächsische Bibliografie	81.358
<input type="checkbox"/> Verbundkatalog Film	25.126
<input type="checkbox"/> Fachkatalog Technik...	22.922
mehr ...	

Erscheinungsjahr ▲	
Von:	Bis:
<input type="text"/>	<input type="text"/>
	
<input type="button" value="Ändern"/>	

Kalkutta : eine moderne Stadt am Ganges



Zugang	Beschreibung	Themen
Personen und Körperschaften:	Waligora, Melitta (Verfasserin) ; Orschied, Rita (Illustratorin) ; Thangarajah, Tilo (Illustratorin)	
Titel:	Kalkutta : eine moderne Stadt am Ganges / Melitta Waligora ; mit Fotografien von Rita Orschied & Tilo Thangarajah	
veröffentlicht:	Berlin : Neofelis Verlag, [2015]	
Umfang:	211 Seiten : Karten ; 285 mm x 245 mm	
ISSN/ISBN:	9783958080140 3958080146	
Format:	Buch , Bildband	
Quelle:	Verbunddaten SWB	
Sprache:	Deutsch	

...?author=Waligora, Melitta

Gesetze des Landes Brandenburg. Ergänzungsbd., Gesetze des Landes Brandenburg : Ergänzungsband



Zugang	Beschreibung
Personen und Körperschaften:	Knöll, Hans-Jochen (HerausgeberIn)
Titel:	Gesetze des Landes Brandenburg. Ergänzungsbd., Gesetze des Landes Brandenburg : Ergänzungsband / hrsg. unter Beratung von Hans-Jochen Knöll
veröffentlicht:	München : Beck, 20XX
Umfang:	Losebl.-Ausg.
Teil von:	Gesetze des Landes Brandenburg.; Ergänzungsbd.
Format:	Buch
Quelle:	Verbunddaten SWB
Sprache:	Deutsch

.../Collection/0014714776

« Vorheriger

Mittellateinisches Wörterbuch bis zum ausgehenden 14. Jhd. Bd. 4



#2 von 250 611 Treffern

- Lateinische Lexika
- Rechtswissenschaft
- Deutsche und ausländische Rechtsge-
schichte
- Allgemeines
- Hilfsmittel und
Nachschlagewerke
- Lateinische Lexika

Zugang	Beschreibung	Thema
RVK-Notation:		
	FR 2872	Ähnliche Treffer finden
	PW 1552	Ähnliche Treffer finden
	NC 2800	Ähnliche Treffer finden

[/Search/Results?lookfor="PW 1552"&type=rvk](/Search/Results?lookfor=)

QUALITÄT EXTERNER QUELLEN

- Gut bis sehr schlecht
- Misst sich an
 - Bezugswege (manuell / automatisierbar)
 - Updateroutinen
 - Datenmodell, Datenformat
 - Vollständigkeit, Richtigkeit, Konsistenz
 - Aktualität
 - Stabilität Links

BEISPIELE SCHLECHTER QUALITÄT

```
<author>fr</author>
```

```
<author>info@maschinenmarkt.de</author>
```

```
<author></author>
```

```
<authors>  
  <author>  
    Möller, Christina  
  </author>  
  <author>  
    Verfasser  
  </author>  
</authors>
```

```
<Publication-Title>
```

```
Das Mensch-Tier-Verhältnis : eine  
sozialwissenschaftliche Einführung,  
Renate Brucker &lt;Hrsg.&gt; ; Frank  
Thieme &lt;Hrsg.&gt; ; Melanie Bujok  
&lt;Hrsg.&gt; ; Birgit Mütherich  
&lt;Hrsg.&gt; ; Martin Seeliger  
&lt;Hrsg.&gt;, Wiesbaden: Springer VS,  
S. 269-297 2015
```

```
</Publication-Title>
```

```
<Title>
```

```
Recruiting im Maschinenbau
```

```
Studenten planen ihre Karriere
```

```
</Title>
```


BEISPIELE SCHLECHTER QUALITÄT

Beispiele für das Feld *Erscheinungsvermerk*

```
! Mirovaja ekonomika i mezdunarodnye otnosenija  
!007338260!Innen- und Aussenpolitik  
!0186286641  
!Gemeinsame Wege ; S. 123-129 : Anm., Lit.  
!Madre mia !  
!nspire!  
"  
" ... einfach weg aus meinem Leben"  
06.12.84  
06/1980 (57 S.)  
06verkusen  
1. Aufl., 1970. - 835 S  
10.
```

Beispiele für das Feld *ISSN*

```
0179-2415Requirements made of artificial ca  
0342-376XR0342-376X  
0720-2679 = IKA. Zeitschrift für internatio  
109I0930-7095  
1615-5742 = Pediatric and developmental pat  
21I0930-7095  
Klischeehafte Sprache  
Anton1439-8001
```

QUALITÄT METRIKEN (AUSWAHL)

- "Uniqueness" des Titels
- Erscheinungsjahr größer als aktuelles Jahr
- Feld-Syntax (z.B. ISSN, URL)
- Feldlängen
- „Excessive Punctuation“
- Pflichtfelder für Linking

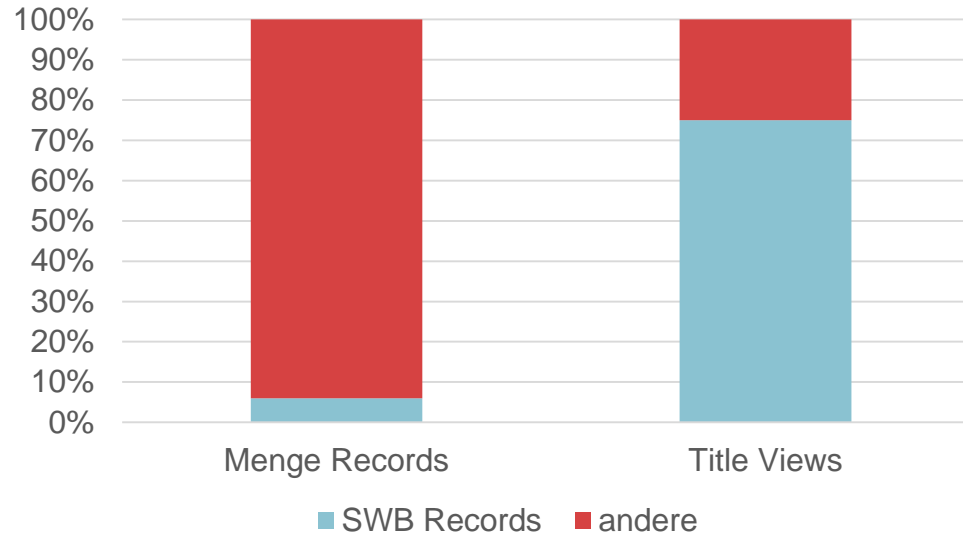
ARBEITSFELD UND HERAUSFORDERUNG

- Metadatenakquise, Rechteklärung
- Bibliothekarisch/Technische Bearbeitung
- Kollektions-Management (Knowledgebase)
- Lizenz-Abgleich (Holdings)
- Linking
- Qualitätssicherung, Fehlerbehebung
- Kommunikation
- = 1 „Metadatenbeauftragte“ (Ref. E-Medien)
- = 1,5 Entwickler / Indexierer

**„...DENNOCH ERFÜLLEN DIE
BIBLIOTHEKSKATALOGE OFFENBAR
WEITERHIN IHREN ZWECK...“**

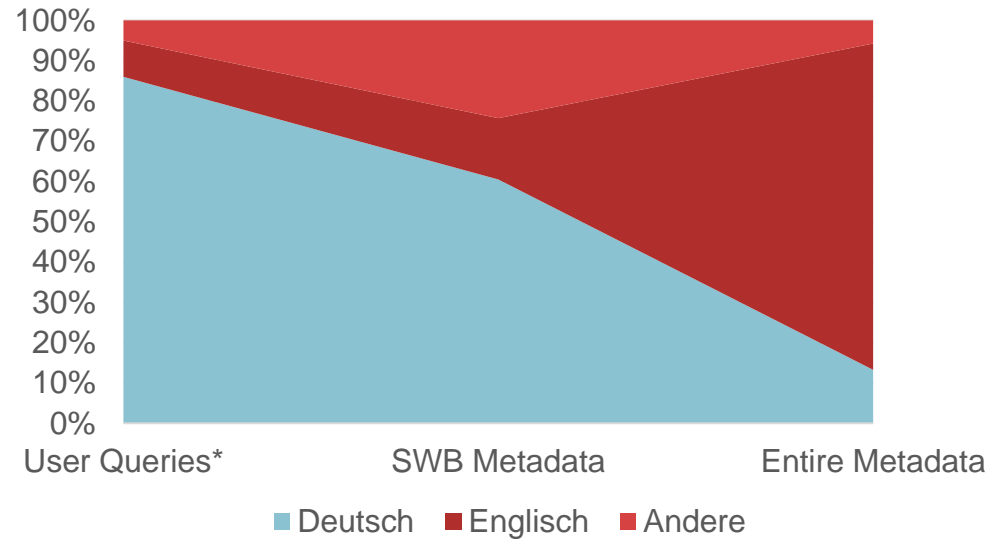
STATISTIK

VERHÄLTNIS MENGE ZU NUTZUNG DER METADATEN



STATISTIK

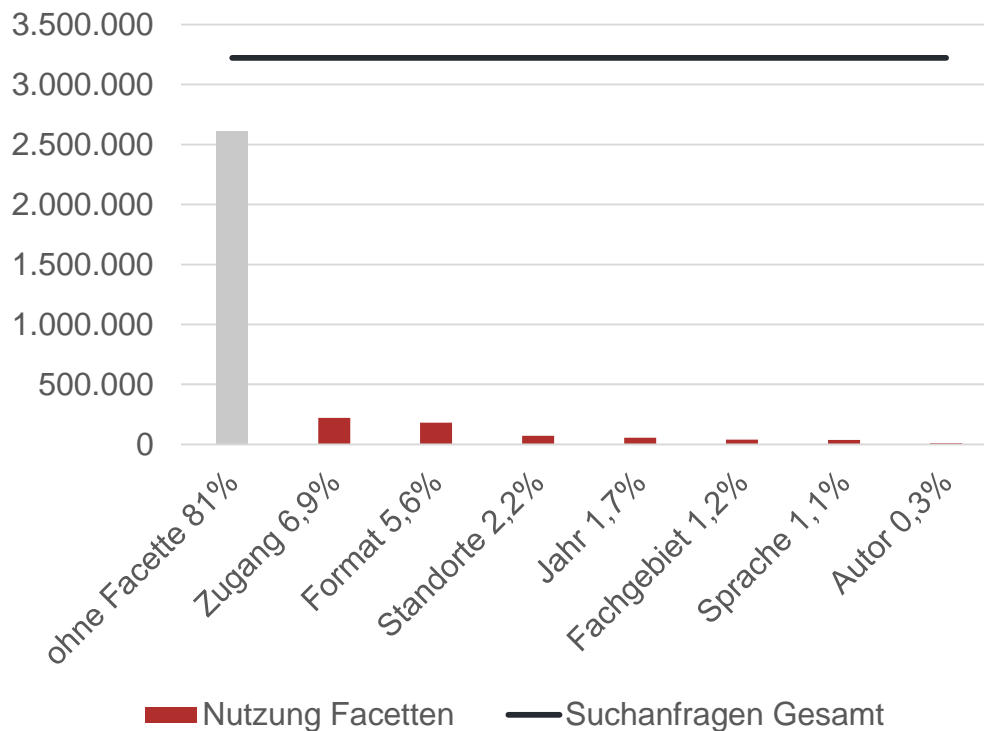
LANGUAGE MATCH



*11/2017

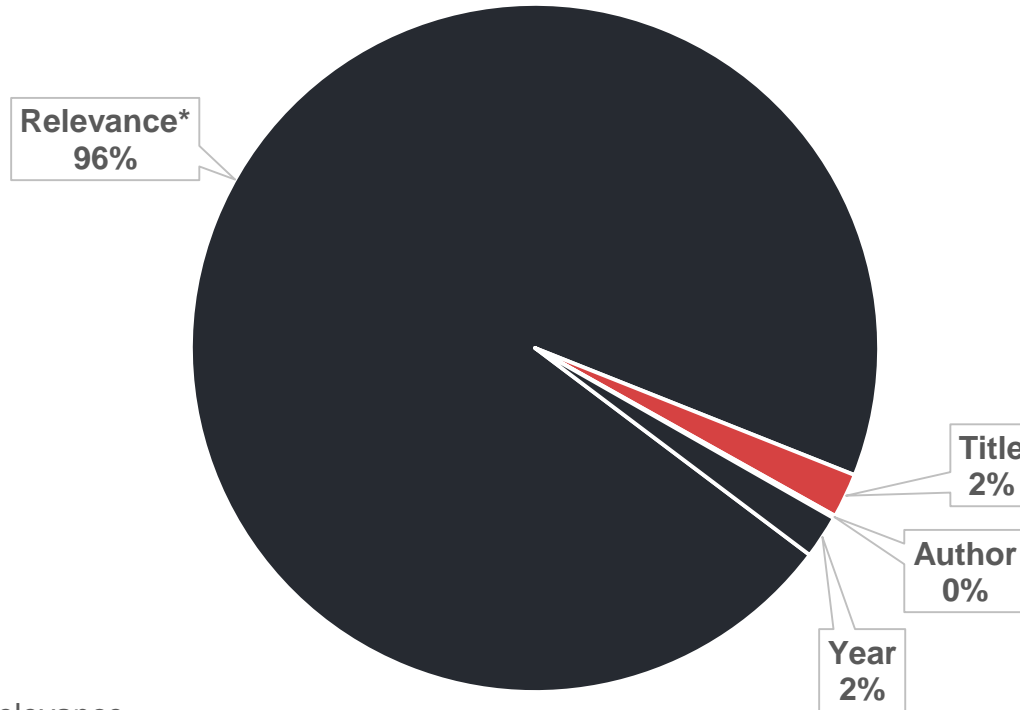
STATISTIK

NUTZUNG FACETTEN



STATISTIK

SORTIERUNG

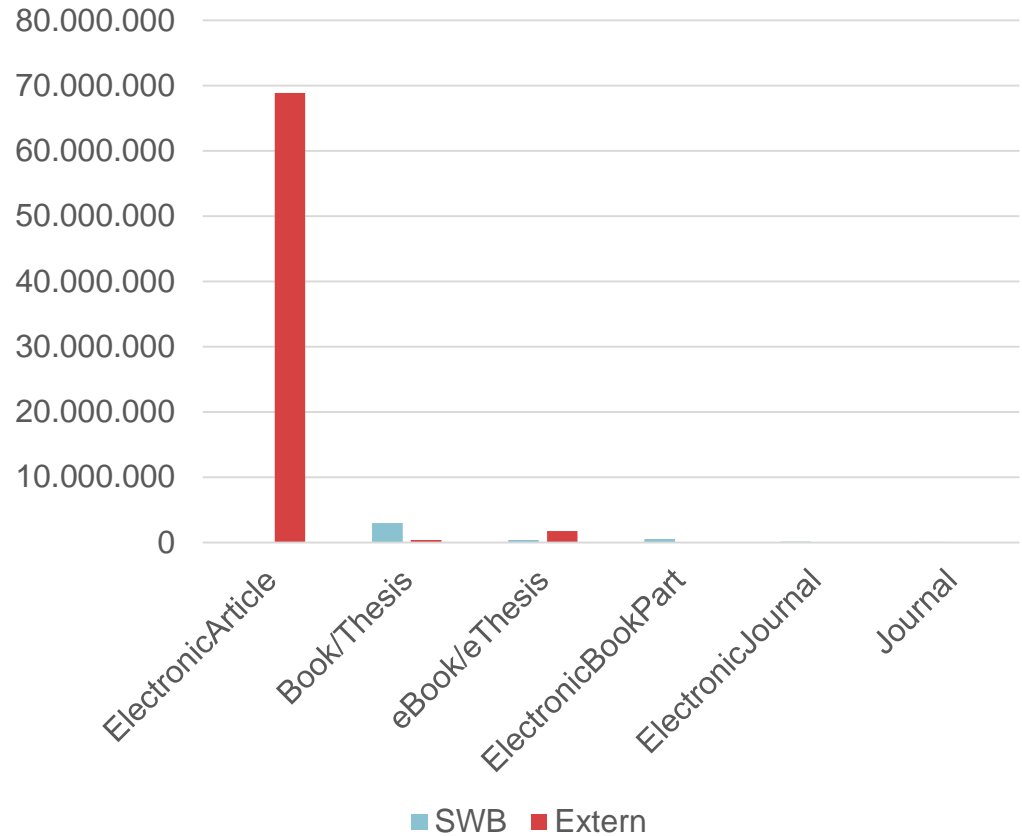


*default sort: Relevance

**„SOLLTE MAN [...] DEN WEG DER AUTOMATISIERUNG UND
DES OUTSOURCINGS DER METADATENERSTELLUNG
KONSEQUENT WEITERGEHEN UND VOLLSTÄNDIG AUF DIE
REGELWERKSKONFORME QUALITÄTSKATALOGISIERUNG
VERZICHTEN?“**

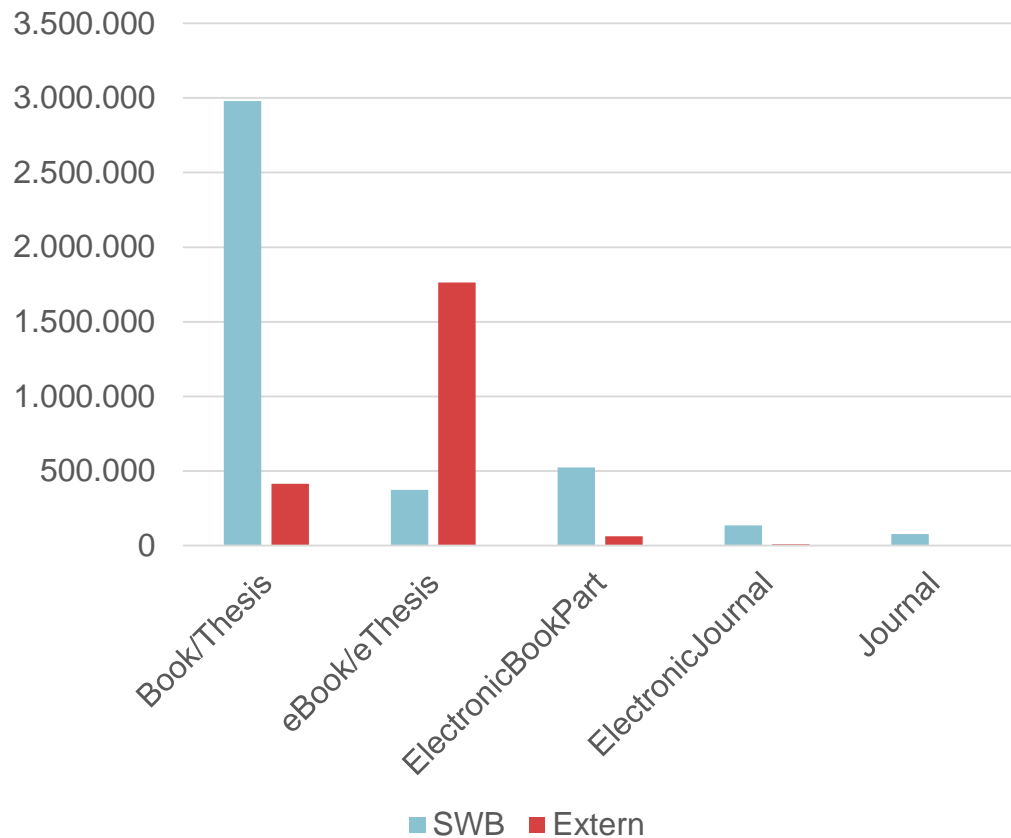
STATISTIK

OUTSOURCING – STATUS QUO



STATISTIK

OUTSOURCING – STATUS QUO
(OHNE E-ARTICLE)



„SOLLTE MAN [...] DEN WEG DER AUTOMATISIERUNG UND DES
OUTSOURCINGS DER METADATENERSTELLUNG
KONSEQUENT WEITERGEHEN **UND VOLLSTÄNDIG AUF DIE
REGELWERKSKONFORME QUALITÄTSKATALOGISIERUNG
VERZICHTEN?**“

ERFASSUNGSSYSTEME

- Verbundsysteme (PICA, ALMA, ZDB, ...)
- Materialspezifische Erfassungssysteme (Portale)
- Institutionelle Repositorien
- Digitalisierungssoftware
 - Strukturdaten
- Knowledgebases (EZB, ... *LAS:eR*, *GoKB*)
 - Paketverwaltung, Lizenzzeiträume
 - Bestandsnachweis, Linking
- Update-Routinen? Aktualität?

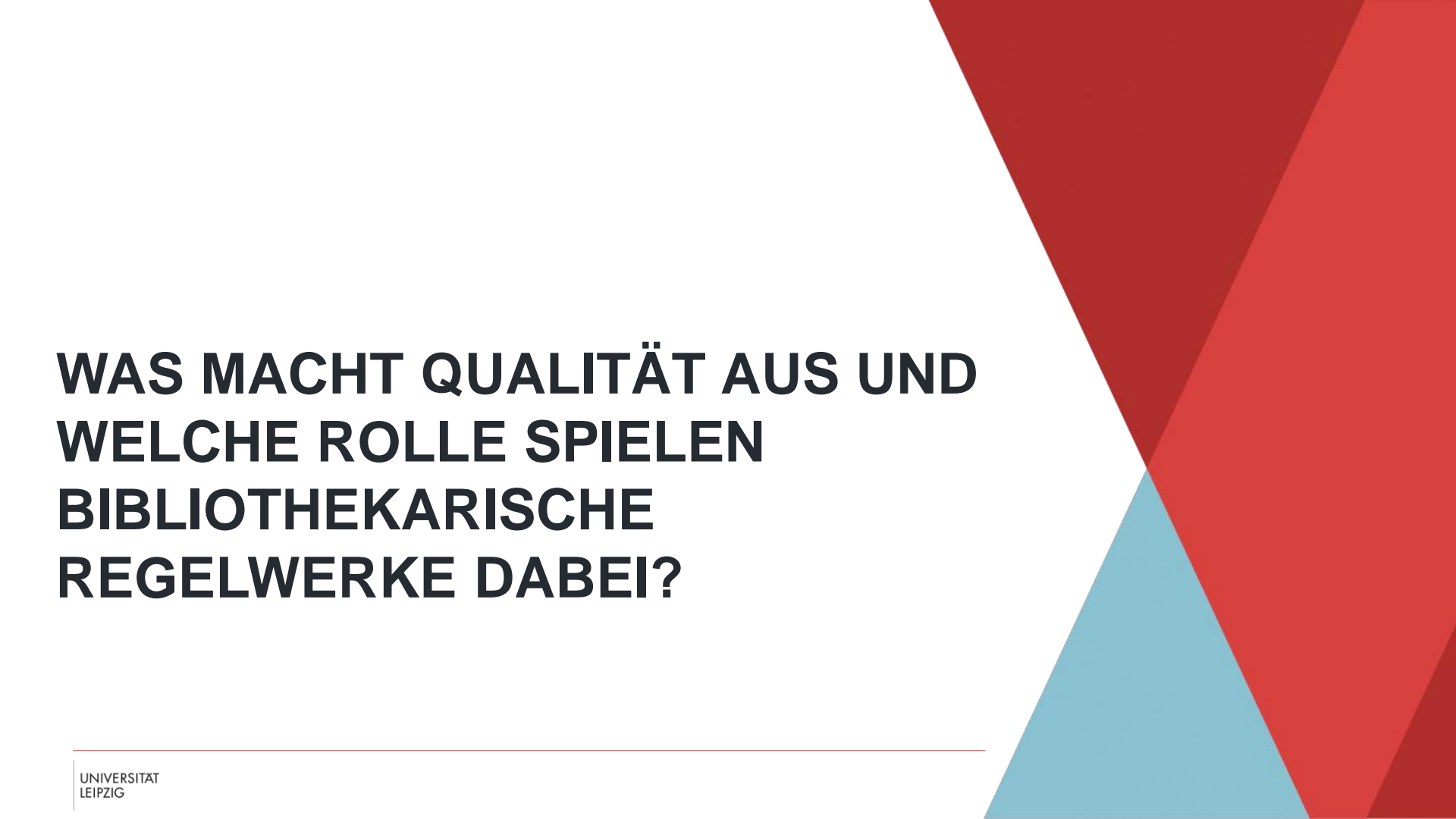
HOMOGENE INFORMATIONSÄÄUME?

- Lokale Digitale Sammlungen?
- Digital Libraries (DDB/Europeana)?
- Institutionelle Repositorien?
- Hochschulbibliographien?
- Forschungsinformationssysteme?

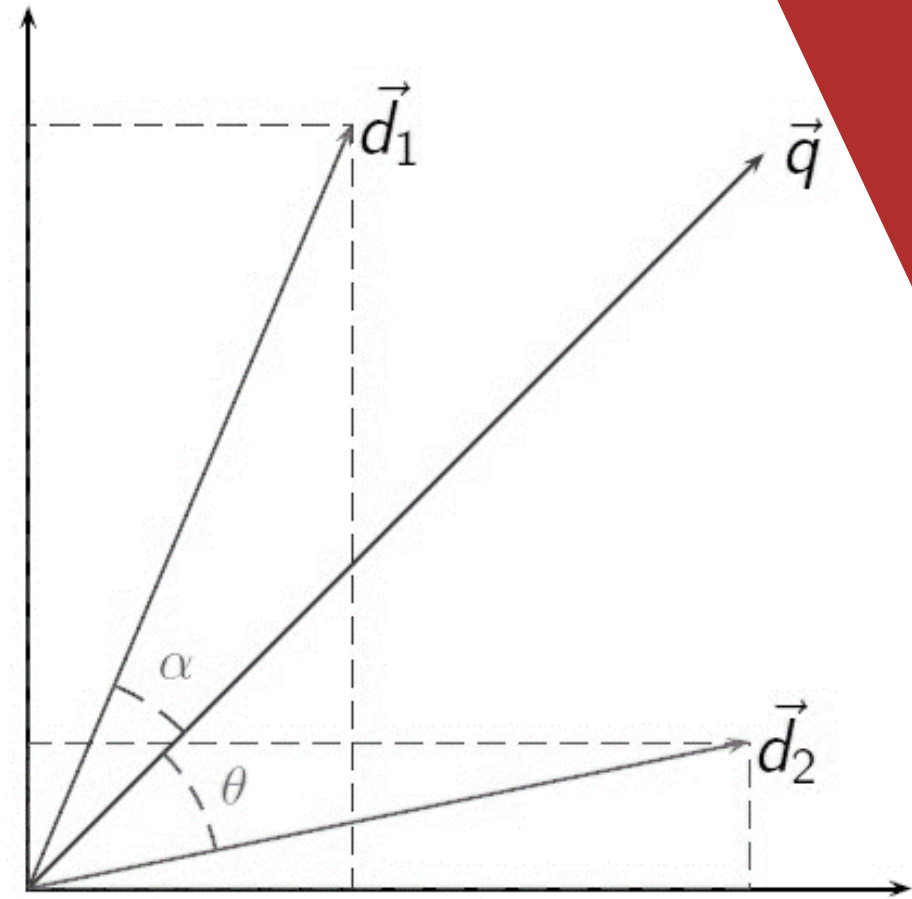
- Besondere Anforderungen?
 - Strukturerschließung?
 - Sacherschließung?
 - Semantische Verknüpfungen?

Welche Qualität brauchen wir?

WELCHE METADATEN MEINEN WIR?



WAS MACHT QUALITÄT AUS UND WELCHE ROLLE SPIELEN BIBLIOTHEKARISCHE REGELWERKE DABEI?



INFORMATION RETRIEVAL

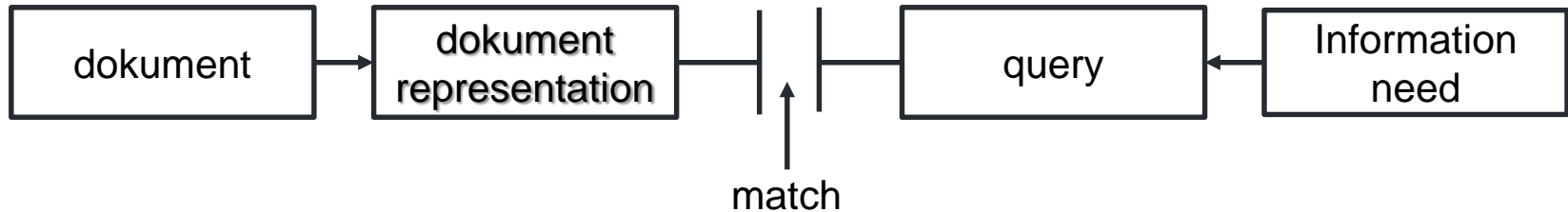
INFORMATION RETRIEVAL

IN BIBLIOTHEKEN

- Vector Space Model (VSM)
- Lookup/String-based retrieval model
- Textstatistische Verfahren für (Voll-)textretrieval
- Term-spezifische Gewichtung
 - Term Frequency-Inverse Document Frequency (Tf-idf)
- Grad der Ähnlichkeit von Anfrage und Dokument
- Partial Term Matching
- Precision und Recall als Qualitätsmetrik

INFORMATION RETRIEVAL

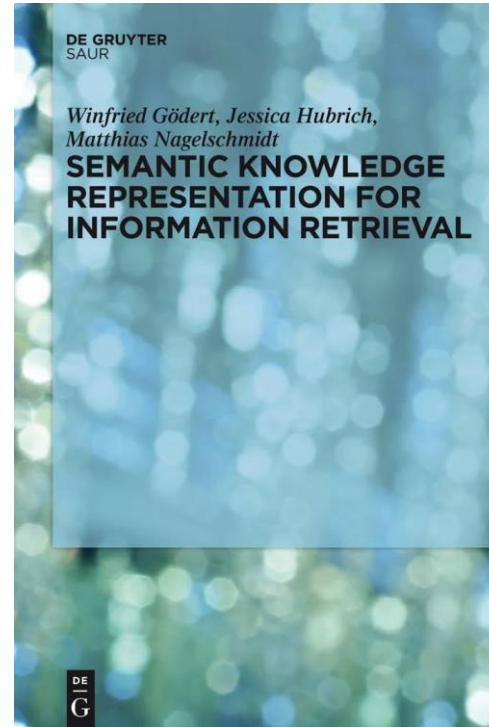
IN BIBLIOTHEKEN



Bates, M.J. (1989). The design of browsing and berrypicking techniques for the online search interface. Online Review , 13(5), 407 – 424.

INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

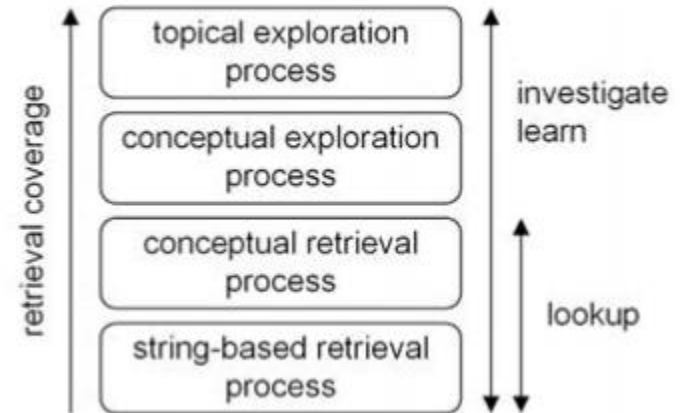
Gödert, W., Hubrich, J., &
Nagelschmidt, M. (2014).
*Semantic knowledge
representation for information
retrieval.*



INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

3 ARTEN VON „INFORMATIONSBEDÜRFNISSEN“ (INFORMATION NEEDS)

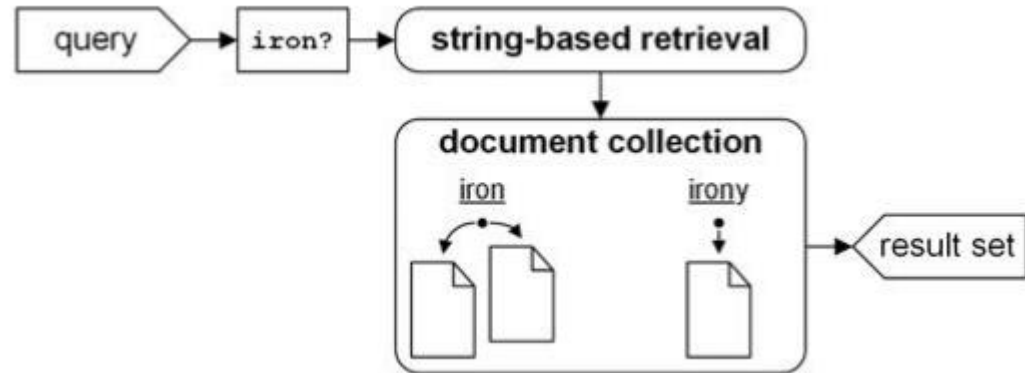
- Lookup
 - known item search
- Learn
 - developing new knowledge
- Investigate
 - analyzing, synthesizing and evaluating extended knowledge



Gödert, W., Hubrich, J., & Nagelschmidt, M. (2014).
Semantic knowledge representation for information
retrieval.

INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

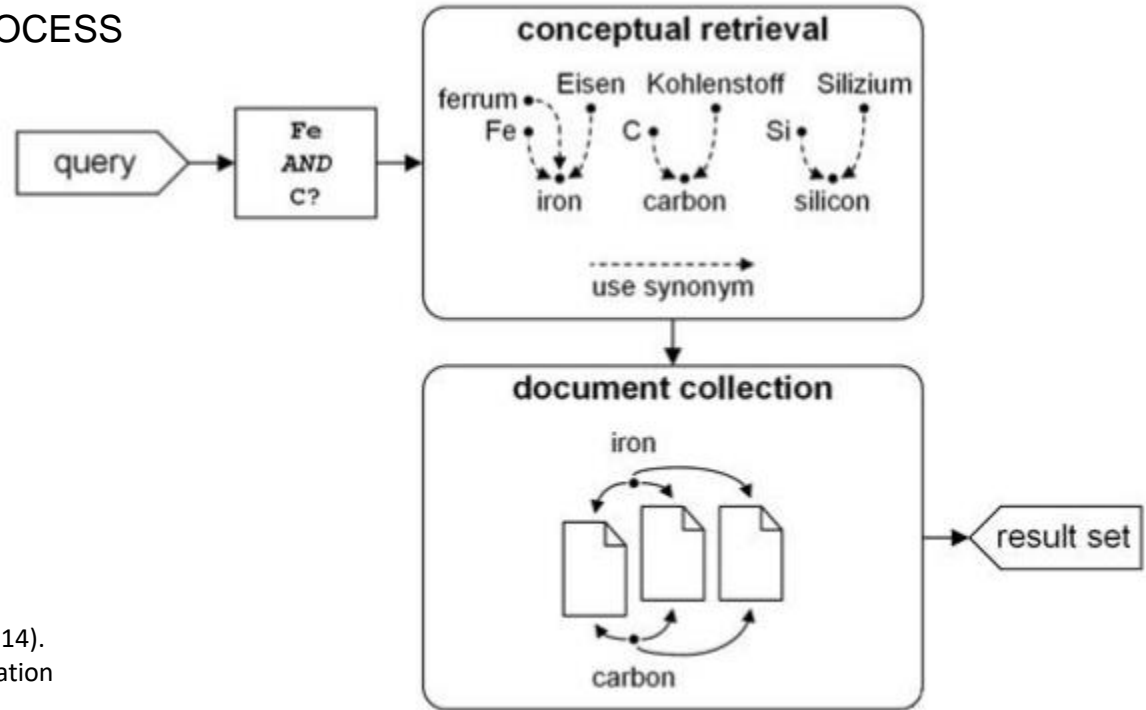
STRING-BASED RETRIEVAL PROCESS



Gödert, W., Hubrich, J., & Nagelschmidt, M. (2014).
Semantic knowledge representation for information
retrieval.

INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

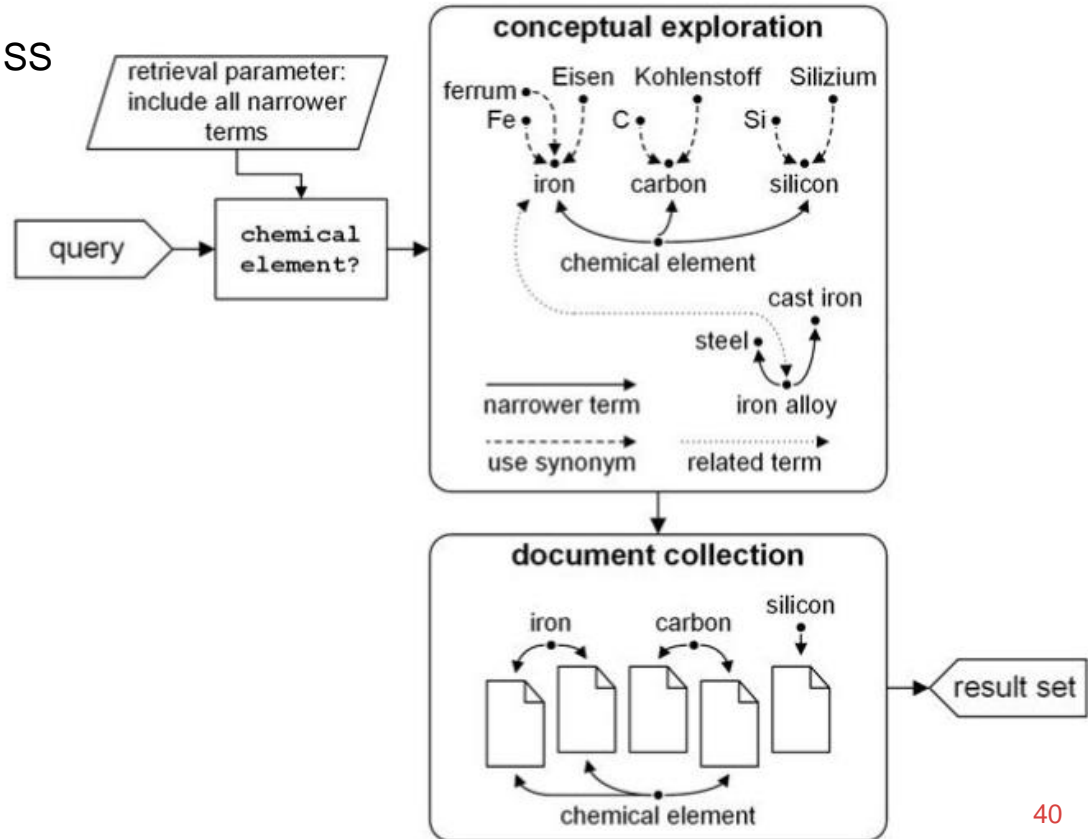
CONCEPTUAL RETRIEVAL PROCESS



Gödert, W., Hubrich, J., & Nagelschmidt, M. (2014). Semantic knowledge representation for information retrieval.

INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

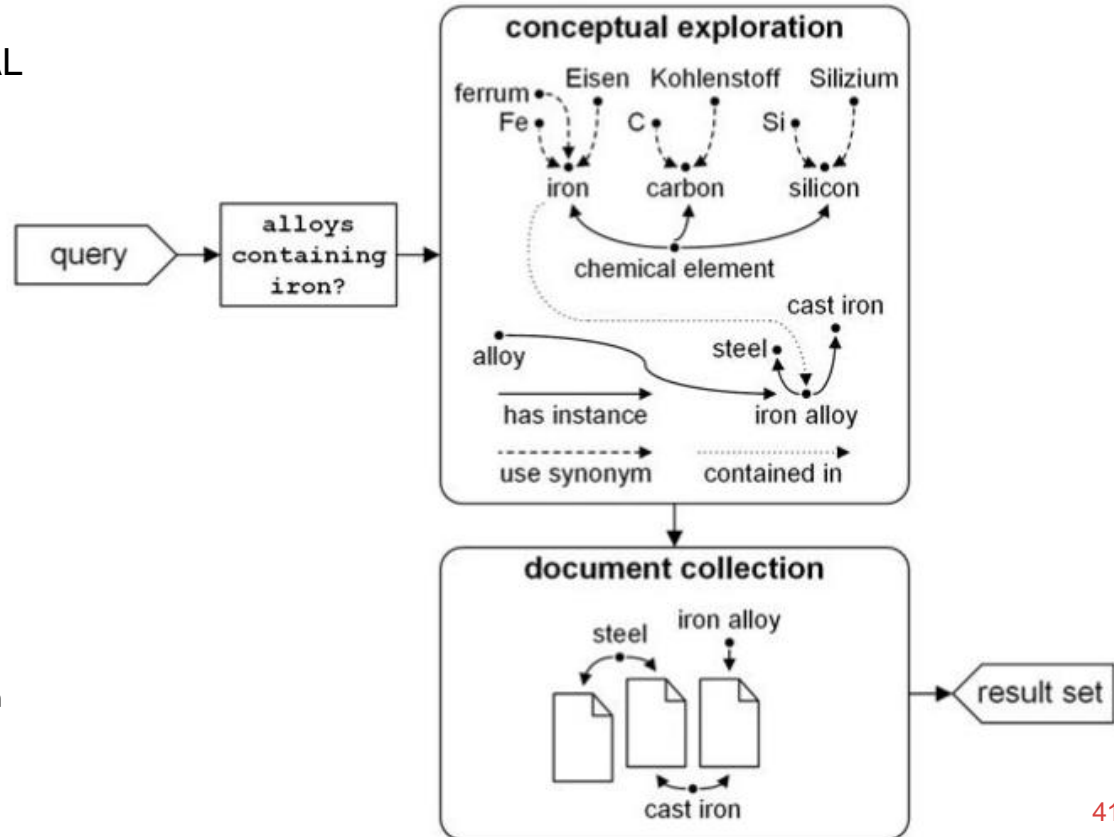
CONCEPTUAL EXPOLARION PROCESS



Gödert, W., Hubrich, J., & Nagelschmidt, M. (2014). Semantic knowledge representation for information retrieval.

INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

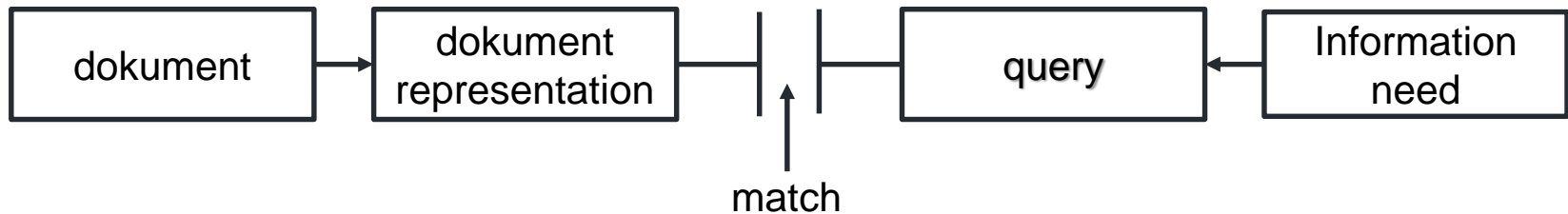
ONTOLOGY-BASED CONCEPTUAL EXPOLARION PROCESS



Gödert, W., Hubrich, J., & Nagelschmidt, M. (2014). Semantic knowledge representation for information retrieval.

INFORMATION RETRIEVAL AND KNOWLEDGE EXPLORATION

- Basiert zu großen Teilen auf „query expansion“ und Modifikation
 - Query re-formulation
- Beinhaltet die Nutzung von „domain ontologies“ (SKOS, ...)



Welche Qualität brauchen wir?

WELCHE ASPEKTE DER REGELWERKSKONFORMEN FORMALERSCHLIEßUNG SPIELEN BEIM IR EINE ROLLE?

FORMALERSCHLIEßUNG UND INFORMATION RETRIEVAL

- Normdaten unterstützen das Auffinden von Entitäten
- differenzierte, vollständige Anzeige von Informationen
- hierarchische Verknüpfungen

- jedoch (noch) kein FRBR

- wenig Auswirkung bezüglich term match

Welche Qualität brauchen wir?

WELCHE ROLLE SPIELT DIE SACHLICHE ERSCHLIEßUNG BEIM IR?

SACHERSCHLIEßUNG UND INFORMATION RETRIEVAL

- hohe Auswirkung auf term match
- konzeptuelle Methoden

- Automatische Sacherschließung
 - bedingt Vorhandensein von Volltexten
 - auf Basis von Metadaten?
 - in welcher Tiefe?
 - Nachnutzbarkeit?

Welche Qualität brauchen wir?

WELCHE ROLLE SPIELT DIE SUCHANFRAGE SELBST IN IR SYSTEMEN?

Information Retrieval zwischen Autopsie und Automatisierung

WELCHE VISIONEN HABEN WIR? WOFÜR?

ANFORDERUNGEN UND BEDARF

DIE ROLLE DER BIBLIOTHEK

IM HINBLICK AUF METADATEN

- Konsument von Metadaten
 - welche wir zum größten Teil nie selbst erschlossen hätten
- Produzent von Metadaten
 - für Objekte, welche wir zugänglich machen
 - welche nicht erschlossen vorliegen
- Anbieter von Content
 - welchen wir kontextuell vernetzen und sichtbar machen wollen
 - digitale Sammlungen für die Forschung
 - Forschungs-Output
- Anbieter von IR Systemen

DIE AUFGABE DER BIBLIOTHEK

IM HINBLICK AUF METADATEN

- Konsument von Metadaten
 - Anbieterkommunikation
- Produzent von Metadaten
 - ...
- Anbieter von Content
 - ...
- Anbieter von IR Systemen
 - ...

METADATA CURATION

- auch für bibliothekarische Metadaten
- „Fremd-Digitalisate“ entsprechend eigenem Print-Bestand einbinden und verknüpfen (Bestandsschutz)



UNIVERSITÄT
LEIPZIG

VIELEN DANK!

Björn Muschall

Universitätsbibliothek Leipzig

Beethovenstraße 6, 04107 Leipzig

T +49 341 97-30508

muschall@ub.uni-leipzig.de

www.ub.uni-leipzig.de