



Erschließung für Discovery-Systeme gestalten

Knowledge Cafe – Netzwerk maschinelle Verfahren in der
Erschließung

Dr.-Ing. Jan Frederik Maas

Zum Vortrag

- Inhalt: Ein Kessel Bunes...
- Grundlage: „**Discovery-Systeme, die ihrem Namen Ehre machen: Verbesserungspotenziale für bibliothekarische Suchmaschinen**“ Keßler et al, Bibliothekskongress 2019
- + Ergänzungen
- **Anregungen** an (Sach-)erschließung aus Sicht der Discovery-Systeme



Discovery als Anwendungsfall

- (Sach-)Erschließung hat viele Anwendungsfälle
- **Erfolgreiches Recherchieren von Werken** zentraler Anwendungsfall
- Problem: Recherche-unterstützende Technik unterliegt regelmäßigen Paradigmenwechseln. Beispiele:
 - Zettelkataloge
 - OPACS
 - Discovery-Systeme
 - ? (Wissensgraphen, semantische Suche...)
- Erschließung und Recherchetechnologie sind oft nicht aufeinander abgestimmt



Discovery als Anwendungsfall

- Zentrale Frage: Wie kann (Sach-)Erschließung das aktuelle Rechercheparadigma (und zukünftige) optimal unterstützen?
 - Anwendungsorientierung
- Von welchen Anwendungsfällen gehen wir bei der Erschließung und der Gestaltung von Erschließung aus? Wie modellieren wir diese?
 - Scenario based-design
 - Personas
 - Storytelling



Discovery: Status Quo

- Indexierung von bibliografischen Metadaten durch Websuchmaschinen-Technologie (SOLR, Elasticsearch)
 - Indexierung von Volltexten / Abstracts / TOCs möglich
 - Relevanzsortierung als Feature
- Oft individuelle Frontends
- Indexschema: VuFind/Marc21 als offener de facto-Standard
 - Metadatenkonversion oft nötig
- Problem heterogen erschlossener Daten
 - Verschiedene Datenquellen
 - Zeitlich variierende Erschließungspraktiken



Alex Saunders / Flickr.com
CC BY-NC.0

Discovery: Anforderungen für SE

- Anwendungsgebiete der Erschließung in Discoverysystemen
 - Finden
 - Relevanzsortierung
 - Darstellung, Vernetzung und Browsing
 - ...?

- Wie sehen „gute Daten“ für Discoverysysteme aus?



Alex Saunders / Flickr.com
CC BY-NC.0

Anwendung: Finden

- Anreicherung der Metadaten durch Schlagworte / Suchbegriffe
 - Je mehr, desto besser (bei hoher Qualität)
 - Unausgesprochene Themen
 - Alternative Ansetzungsformen
 - Ansetzungsformen widerspiegeln Wortwahl der Recherchierenden(?)
 - Paradigmatische Relationen (= Unterbegriffe, Oberbegriffe) nützlich?
- Facettierung
 - Hohe Abdeckung
 - Sprechende Namen pro Kategorie
 - BK deutlich besser als RVK
 - Eindeutigkeit im Themenbaum

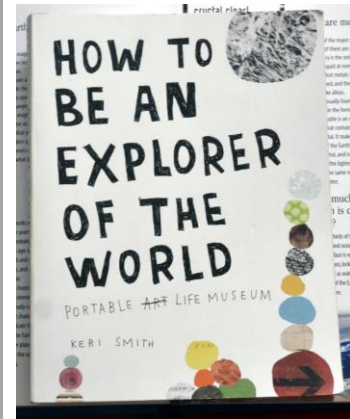


INDEX CARD

Justin See/ Flickr.com
CC-BY 2.0

Anwendung: Relevanzsortierung

- Relevanzranking = Algorithmus der Gewichtung über verschiedenen Feldern der Datensätze
- Voraussetzung: Maschineninterpretierbare Felder
 - Beispiel: Auflagenkennzeichnung
 - Auswahl der Felder mit direkter Auswirkung auf Relevanz
- Problem: Datenheterogenität aufgrund wechselnder Erschließungspraktiken
 - Beispiel: Bevorzugung umfangreicher erschlossener Metadatensätze durch TF gegenüber aktuellerer Literatur
- Ranking durch machine learning-Verfahren
 - Optimierung des Rankings durch spezielle Sacherschließungsmerkmale / Zuarbeit aus der Sacherschließung?



Anwendungen: Darstellung, Vernetzung, Browsing

- Discovery-Systeme müssen die Metadaten-Komplexität nachmodellieren, um diese nutzen zu können
 - Beispiel: Verknüpfung mit GND, Systematiken
 - IT-Ressourcenproblem
 - Generisches Erfassen/Darstellen von Relationen über Erschließungssysteme hinweg -> sinnvoll/machbar?
- Keine Optimierung der Feldbefüllung für spezifische Frontends
- Auch hier: Maschineninterpretierbare Felder



Fazit

- Wandel des Rechercheparadigmas berücksichtigen und gestalten
- Konstruktiver Dialog über Gegenwart und Zukunft der Recherche
 - Fachtagung :)
- Dazu kooperativer Einsatz von Techniken zur Konkretisierung der Anwendungsszenarien sinnvoll?



Vielen Dank

Dr.-Ing. Jan Frederik Maas

Von-Melle-Park 3
20146 Hamburg

040 / 4 28 38-66 74
jan.maas@sub.uni-hamburg.de

 www.sub.uni-hamburg.de
facebook.com/stabihh
 twitter.com/stabihh