

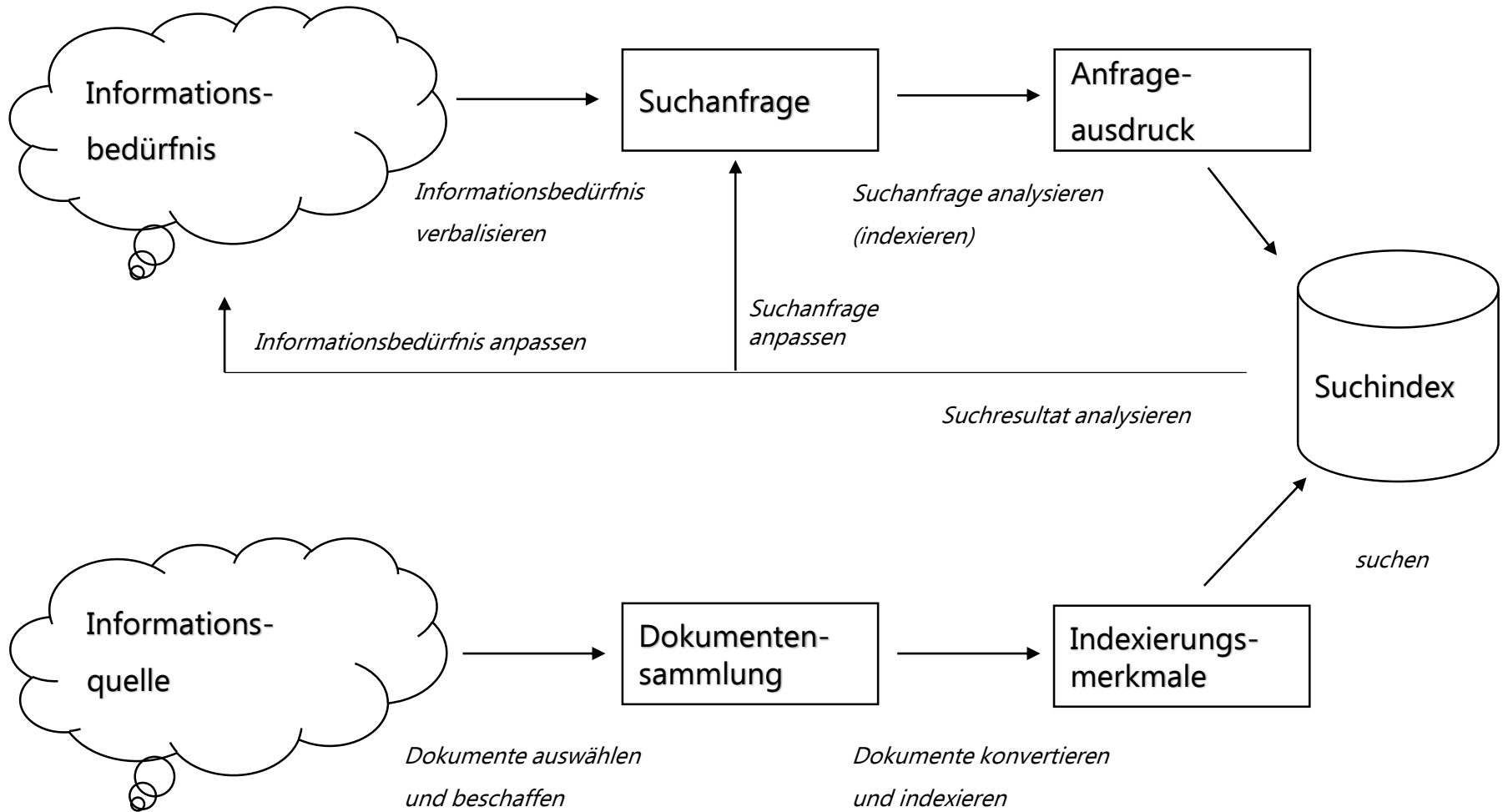
BibSpider

*Eine neue Recherchemethode für
Schlagwort- und Titelsuche*

05.09.2018, Adriano Meyer Broyn



- Zusammenspiel von Titeln und Inhaltsererschließung besser verstehen.
 - Formal und inhaltlich ähnlich Publikationen sollten inhaltlich ähnlich erschlossen sein (van Rijsbergen Cluster-Hypothese).
 - Grundlagen für maschinelles Lernen.
- Prototyp für eine neue Ähnlichkeitssuche in DA-3
 - Im DA-2 wurden nur Beziehungen zwischen Titeln berücksichtigt, nicht zwischen Schlagwörtern und Titeln.
 - Nutzung von approximativen Übersetzungen.
- Die Sacherschließung für die Benutzer besser nutzbar machen
- Langfristig: Teilautomatisierte Erschließung.
 - Bedingt automatische, zuverlässige Schätzung der Qualität. Maschine sollte selbst herausfinden, ob berechnetes Erschließungsergebnis gut ist oder nicht.
 - Probabilistisches Anordnungsprinzip ist i. Allg. wenig hilfreich

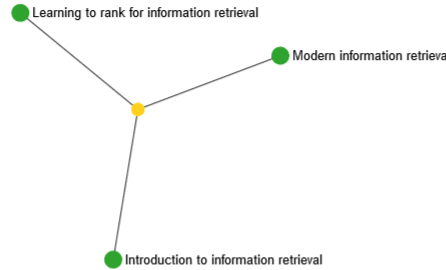


- Webinterface um nach Titel und Schlagwörter zu suchen.
- Benutzer wählt einige relevante Titel und Schlagwörter aus.
- BibSpider berechnet Vorschläge für weitere Titel und Schlagwörter.
- Die aktuelle Auswahl inkl. Verknüpfungen wird graphisch dargestellt.
- Es können jeder Zeit weitere Informationen zu den Titel oder Schlagwörter angezeigt werden.

<http://bibspider-test.eurospider.com/>

- Wir suchen das Schlagwort «Information Retrieval».
- Es werden Bücher, die mit diesem Schlagwort erschlossen worden sind, vorgeschlagen.
- Wählen dann die Bücher:
 - «Learning to rank for information retrieval»
 - «Modern information retrieval»
 - «Introduction to information retrieval»

BibSpider - Prototyp



Introduction to information retrieval
 ID: z01_978-0-521-86571-5_01
 Subtitle:
 Author: Manning, Christopher D.
 ISBN: 978-0-521-86571-5 (hardback)
 Source: zbz
 LCSH: [650] Document clustering,[650] Information organization,[650] Information retrieval,[650] Information storage and retrieval systems,[650] Semantic Web,[650] Text processing (Computer science)
 RVK:
 DDC: 025.04
 GND:
 4059667-9: Textverarbeitung
 4688372-1: Semantic Web
 4072803-1: Information Retrieval

Gewähltes Schlagwort

Schlagwort suchen ... 🔍 ✕ Select

- Information Retrieval info
- World Wide Web info
- Datenbanksystem info
- Multimedia info
- Data Mining info
- Ontologie <Wissensverarbeitung> info
- Abfrage info
- Content Management info
- Bildbanksystem info
- Semantic Web info
- Informationssystem info
- Suchmaschine info
- Information-Retrieval-System info
- Dokumentverarbeitung info
- Wissensextraktion info
- Natürlichsprachiges System info

Buch suchen ... 🔍 ✕ Select

- Learning to rank for information re... info
- Modern information retrieval info
- Introduction to information retriev... info
- Modern information retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information retrieval technology info
- Information retrieval technology info
- Information retrieval technology info
- Information retrieval technology info
- Information retrieval technology info
- Human information retrieval info
- Information retrieval technology info

Vorgeschlagene Schlagwörter

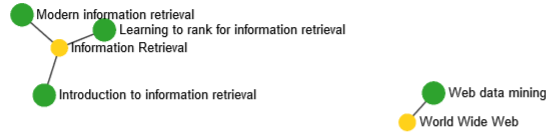
Vorgeschlagene Bücher

Gewählte Bücher

- Als Schlagwort wird uns als aller erstes «World Wide Web» empfohlen.
- Wir wählen «World Wide Web» aus.
- Es werden uns neue Bücher vorgeschlagen die sowohl mit «information retrieval» als auch «World Wide Web zu tun haben»
- Die Resultate sind auch Sprachunabhängig: Mit einem englischen Schlagwort wurden uns auch deutsche oder französischen Bücher vorgeschlagen.
- Der Graph dient als Übersicht für die ausgewählten Schlagwörter und Bücher. Der Benutzer kann damit schnell nachschauen welche Schlagwörter auf welche Bücher zeigen.



BibSpider - Prototyp



Schlagwort suchen ...

- Information Retrieval info
- World Wide Web info
- Anwendungssystem info
- Data Mining info
- Datenbanksystem info
- Informationssystem info
- Suchmaschine info
- Multimedia info
- Semantic Web info
- Ontologie <Wissensverarbeitung> info
- Wissensextraktion info
- Abfrage info
- Computersicherheit info
- Benutzerorientierung info
- Benutzermodell info
- Content Management info

Buch suchen ...

- Learning to rank for information re... info
- Modern information retrieval info
- Introduction to information retriev... info
- Web data mining info
- Recherche d'information info
- Information retrieval info
- Information Retrieval info
- Dynamic taxonomies and faceted sear... info
- Modern information retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Information Retrieval info
- Web Data Mining info

FR-Vorschlag

DE-Vorschlag

Information retrieval

ID: swb-103734961

Subtitle: Suchmodelle und Data-Mining-Verfahren für Textsammlungen und das Web

Author: Ferber, Reginald

ISBN: 3898642135

Source: swb

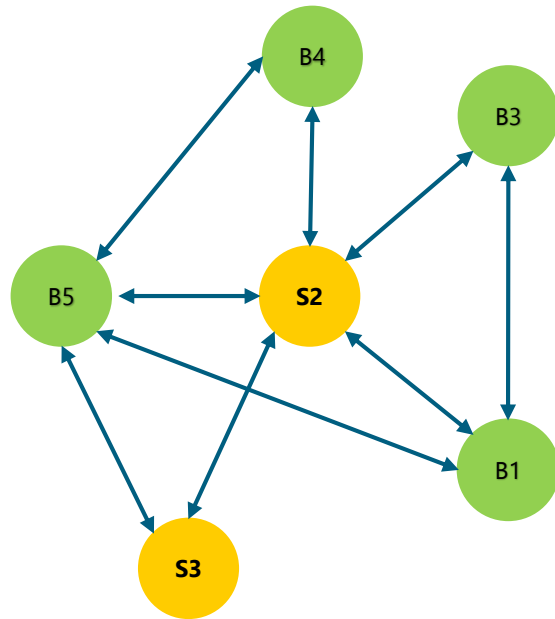
LCSH:

RVK: AN 95000,QH 500,ST 270,ST 530

DDC:

GND:

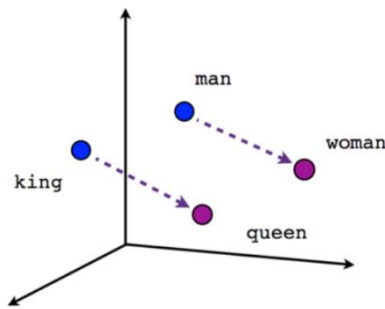
- 4363898-3: World Wide Web
- 4428654-5: Data Mining
- 4072803-1: Information Retrieval



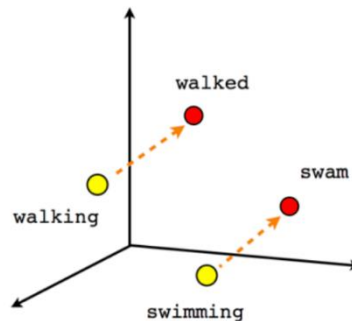
- Die Ähnlichkeit kann als die gegenseitige Anziehungskraft zwischen zwei Massenpunkte betrachtet werden.
- Die Federn haben unterschiedlichen Stärke:
 - Stark: inhaltliche grosse Abweichung.
 - Weich: inhaltliche grosse Übereinstimmung.

Möglichkeiten die semantische Nähe von Schlagwörtern zu berechnen:

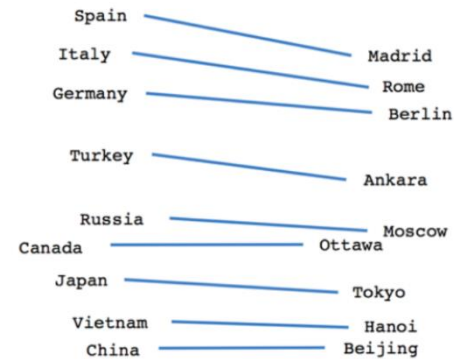
- Konkordanzen: Die Schlagwörter eines Titels aus einem anderen Beschlagwortungssystem (LCSH, RVK, DDC) für die Erschliessung einberechnen.
- Co-occurrence: Wenn zwei Schlagwörter sehr oft zusammen vorkommen, dann sind sie wahrscheinlich auch semantisch ähnlich.
- Word2vec:



Male-Female



Verb tense



Country-Capital

- Performance:
 - Strukturierung und Reinigung der Daten.
 - Sinnvolle Indexierung
- Datensatz: Mit einer grösseren Datenmenge können die Ähnlichkeitsfunktionen besser angewandt werden.
- Machine Learning um Ähnlichkeitseigenschaften zu trainieren.
- Benutzerfreundlichkeit: BibSpider soll praktisch, schnell und flexibel sein.