Carina Fiedler

### INDEXING

#### Inhalt

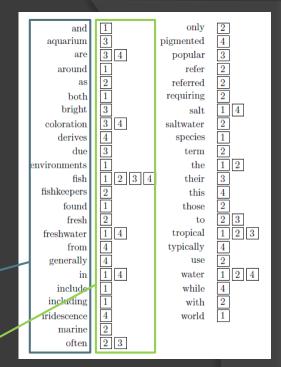
- Motivation Wozu braucht man Indizes?
- Invertierter Index
- Programmierung eines invertierten Index
- Suchanfragen mit Hilfe eines Index

# Motivation – Wozu braucht man Indizes?

- Grundlage: kurze Suchanfragen,
   Elemente nicht in allen Dokumenten vorhanden
- Ohne Index: sequentielles Durchsuchen aller Dokumente => sehr hoher Zeit- und Rechenleistungsbedarf
- Index optimiert Geschwindigkeit und Durchführung, da nur der Index durchsucht werden muss

#### Invertierter Index

- Bekannteste Datenstruktur für Textsuche
- Aufbau:



	l .
Lexikon/Wörterbuch/Vokabular	Autreten/Postings
<ul> <li>Liste aller vorkommenden Terme</li> <li>Zeiger auf die Postings- Struktur</li> <li>Datenstruktur: Hashtabelle, Baumstrukturen</li> </ul>	<ul> <li>Liste aller Positionen, an denen das Wort in den Dokumenten auftritt</li> <li>Dokumenten ID, Anzahl der Dokumente, in denen der Term auftritt, Position im Dokument</li> </ul>

### Programmierung eines Index

- Sammlung der Indexterme
- Alphabetische Sortierung der Terme und docID
- Gruppierung von gleichen Termen
- Aufspaltung in Wörterbuch und Postings
- Sortierung der Postings nach docID

# Suchanfragen mit Hilfe eines Index

#### Invertierter Suchalgorithmus:

- Finde gesuchte Elemente (Wörter, Muster etc.) im Lexikon
- Frage die Auftrittspositionen für jeden Lexikoneintrag ab
- Bewerte die Relevanz der Auftrittspositionen bzw. der Dokumente entsprechend Retrieval Modell

#### Quellen

- en.wikipedia.org/wiki/Search\_engine\_indexing
- <u>en.wikipedia.org/wiki/Inverted\_index</u>
- en.wikipedia.org/wiki/Database\_index
- nlp.stanford.edu/IR-book/html/htmledition/a-firsttake-at-building-an-inverted-index-1.html
- www.csee.umbc.edu/~ian/irF02/lectures/04Search
   -Inverted-Files.pdf