

Übungen zu Internet-Suchmaschinen, Sommersemester 2015

Vu Tran (LF 139)

Sprechstunde nach Vereinbarung

vtran@is.inf.uni-due.de

Übungsblatt 2

keine Abgabe

Aufgabe 3: Trunkation und Maskierung

Ein IR-System unterstütze Freitextsuche mittels Trunkierung, Maskierung und Kontextoperatoren wie im Skript bzw. in den Folien vorgestellt.

- (a) Finde Suchanfragen mit Trunkierungs- und Maskierungsoperatoren, die die fettgeschriebenen Terme in den folgenden Listen matchen, nicht aber die nicht-fett geschriebenen Terme. Nimm an, dass die Anfrage komplette Terme matchen muss (also es nicht reicht, wenn das Suchmuster auf einen Teil eines Terms zutrifft).

Die Terme sind:

- **laufen**
- **lief**
- **gelaufen**
- **läuft**
- Luft
- Licht

und

- **Türen**
- **Tür**
- Trier
- Treppe
- Tor
- Türkei
- Türkeil

und

- **Stadt**
- **Städte**
- **Stadtrat**

- Student

- (b) Ein Student soll für ein Seminar ein Referat über die Informationsbedürfnisse mobiler Benutzer halten. Zum Einstieg in das Thema überlegt sich der Student, mittels des oben beschriebenen IR-Systems deutschsprachige Forschungsveröffentlichungen zum Thema zu finden. Dazu darf er in einem Retrievalsystem genau eine Suchanfrage absetzen, um in Volltexten zu suchen. Welche Suchanfrage würdest Du ihm vorschlagen? Das System versteht nur Terme (also keine booleschen Operatoren) und Maskierung, Trunkierung und Kontextoperatoren. Welche Abwägung muss hier getroffen werden?

Aufgabe 4: Head-Modifier-Strukturen

- (a) Finde eine Head-Modifier-Struktur für den Begriff „(Die) Reihenfolge von Ereignissen in verteilten Computersystemen“. Trenne dafür Komposita in ihre Bestandteile.
- (b) Finde eine Nominalphrase zu der folgenden Head-Modifier-Struktur:
(((interactive, (retrieval, interfaces)), examination),
(user, study))
- (c) Gegeben sei eine Anfrage als Head-Modifier-Struktur (knowledge, representation). Welche der folgenden Head-Modifier-Strukturen passen auf die Anfrage?
- (i) (((computer, generated), knowledge), representation)
 - (ii) ((knowledge, based), (strategy, representation))
 - (iii) ((strategy, representation), (new, knowledge))

Aufgabe 5: Feature-Beratung

In das IR-System, das in Eurer Firma verwendet wird, soll ein neues Feature eingebaut werden. Der damit beauftragte Dienstleister hat eine Liste mit möglichen Erweiterungen vorgelegt, die Du hinsichtlich ihrer Auswirkung auf Recall und Precision beurteilen sollst. Die möglichen Erweiterungen sind:

- Stammformreduktion: Eine Suche nach „laufen“ liefert auch Dokumente, die „lief“, „Läufer“, „Lauf“ etc. enthalten
- Thesaurus zur automatischen Erweiterung von Anfragen um Synonyme: Eine Suche nach „Tastatur“ liefert auch Dokumente, die „Keyboard“ enthalten.
- Disambiguierung von Autorennamen in der Form, dass man zwei Autoren mit demselben Namen unterscheiden kann
- Kompositazerlegung: Eine Suche nach „Dampfschiff“ liefert auch Dokumente, in denen „Dampfschiffkapitän“, „Donaudampfschiff“ oder „Donaudampfschiffkapitän“ vorkommt.

- Phrasensuche: eine Suche nach „information retrieval“ liefert nur Dokumente, in denen die beiden Terme in direkter Folge auftreten – also z.B. nicht „retrieval of information“ oder „information about retrieval engines“.

Beschreibe jeweils kurz, in welche Richtung Recall und Precision jeweils nach Integration des jeweiligen Features tendieren würden und wieso.