

## Übungen zu Internet-Suchmaschinen, Sommersemester 2015

Vu Tran (LF 139)

Sprechstunde nach Vereinbarung

vtran@is.inf.uni-due.de

### Übungsblatt 5

keine Abgabe

---

#### Aufgabe 11: Probabilistisches Retrieval – BIR-Modell

Gegeben seien die folgenden Dokumente mit den Termen  $x_1$  und  $x_2$  und den Relevanzbeurteilungen  $r(d_m)$ <sup>1</sup> für eine Anfrage  $q^T = \{x_1, x_2\}$ :

$d_m$	$r(d_m)$	$x_1$	$x_2$
$d_1$	R	1	1
$d_2$	R	1	1
$d_3$	R	1	1
$d_4$	R	1	1
$d_5$	N	1	1
$d_6$	N	1	1
$d_7$	N	1	1
$d_8$	R	1	1
$d_9$	N	1	0
$d_{10}$	N	1	0
$d_{11}$	R	1	0
$d_{12}$	R	1	0
$d_{13}$	R	1	0
$d_{14}$	R	1	0
$d_{15}$	N	0	1
$d_{16}$	R	0	1
$d_{17}$	R	0	1
$d_{18}$	N	0	1
$d_{19}$	N	0	0
$d_{20}$	R	0	0

- (a) Berechne die Werte von  $P(R|d_m^T)$  (ggf. über den Zwischenschritt  $O(R|d_m^T)$ )
- (i) mittels BIR-Modell mit Abschätzungen:  $s_i = \frac{n_i}{N}$  und  $p_i = 0,5$  (siehe Skript).  $O(R)$  kann als 1 gesetzt werden.
  - (ii) mittels BIR-Modell und anhand der Termfrequenzen und der Relevanzbewertungen in o.g. Tabelle
  - (iii) direkt
- (b) Wie sind die Unterschiede zu erklären?

---

<sup>1</sup>„R“ für „relevant“, „N“ für „nicht relevant“

Beachte, dass unterschiedliche Dokumente dieselbe Repräsentation haben können!