

Übungen zu Internet-Suchmaschinen, Sommersemester 2015

Vu Tran (LF 139)

Sprechstunde nach Vereinbarung

vtran@is.inf.uni-due.de

Übungsblatt 9

keine Abgabe

Aufgabe 18: Learning to Rank als Klassifikationsproblem

- (a) Formuliere LTR als Klassifikationsproblem, indem Du eine Tabelle mit Beispieltrainingsdaten für den punktweisen Ansatz¹ bestehend aus Merkmalen und der Relevanzbeurteilung angibst.

Beispiel (für den Scoring-Ansatz):

	q	$m1$
#1	a b	0,4	0,8
#2	a b	0,634	0,015
#3	c	0,122	0,4
...

- (b) Wandele die Tabelle aus Aufgabenteil (a) in Beispieltrainingsdaten für den paarweisen Ansatz² um.
- (c) Welche Probleme bzw. Herausforderungen können beim Einsatz von LTR in der Praxis auftreten? Welche Vor- und Nachteile bringt der Einsatz von LTR?

Aufgabe 19: Diversity ranking

- (a) Für die Anfrage „NSA“ ergeben sich Wahrscheinlichkeiten wie in der Tabelle angegeben. Fülle die Tabelle mit den gesuchten Relevanzwahrscheinlichkeiten $P(R|d_j, q)$ aus. Nimm λ als 0,5 an.
- (b) Wie kann eine Suchmaschine die $P(d_j|q_i)$ schätzen?

„NSA“	$P(d_j q)$	$P(d_j q_i)$				$P(R d_j, q)$
		Snowden	BND	Merkel	Jobs	
$P(q_i q)$		0.7	0.1	0.05	0.15	
d_1	0.9	0.3				
d_2	0.8	0.7	0.3			
d_3	0.7	0.1	0.6			
d_4	0.6		0.4	0.6		
d_5	0.5				0.2	
d_6	0.5				0.4	

¹Retrieval Status Value (RSV) wird vorhergesagt

²Reihenfolge der Dokumente wird vorhersagt (ordinale Klassifikation)