



Informationsintegration

Übung 5

SS 2016

Yvonne Lichtblau

Vorstellung Lösungen Übung 5

Wettbewerb

Gruppe	1	2	3	4	5	6	8
Web Scraping (Korrektheit)	3			1			5
Web Scarping (Geschwindigkeit)	5	3			1		
Hierarchical Queries	3		1				5
Query Containment			5			3	1
Ontology Matching							
Summe	11	3	6	1	1	3	11

Auswertung Übung 5

	ansatz	length	precision	recall	fscore	recall+
Gruppe3	3-gramm-Überschneidung	1373	0.93	0.85	0.8875	0.59
Gruppe2	Tokenmengen/2-gramme	1098	0.99	0.72	0.8332	0.25
Gruppe1	TFIDF basiert	1118	0.95	0.74	0.8319	0.31
Gruppe8	TFIDF basiert	1177	0.93	0.72	0.8162	0.27
Gruppe5	Jaccard-Koeffizient	1213	0.92	0.73	0.8142	0.29
Gruppe4	Jaccard/n-gramme?	1158	0.94	0.72	0.8115	0.24
Gruppe6	Fuzzy String matching module python	1127	0.95	0.70	0.8067	0.21

Wettbewerb

Gruppe	1	2	3	4	5	6	8
Web Scraping (Korrektheit)	3			1			5
Web Scraping (Geschwindigkeit)	5	3			1		
Hierarchical Queries	3		1				5
Query Containment			5			3	1
Ontology Matching	1	3	5				
Summe	12	6	11	1	1	3	11