

Optimizing Search Engines using Clickthrough Data

Thorsten Joachims, KDD'02

- What did it learn from the data?

- Linear correlations

Positive
correlated
features



Negative
correlated
features



weight	feature
0.60	query_abstract_cosine
0.48	top10_google
0.24	query_url_cosine
0.24	top1count_1
0.24	top10_msnsearch
0.22	host_citeseer
0.21	domain_nec
0.19	top10count_3
0.17	top1_google
0.17	country_de
...	
0.16	abstract_contains_home
0.16	top1_hotbot
...	
0.14	domain_name_in_query
...	
-0.13	domain_tu-bs
-0.15	country_fi
-0.16	top50count_4
-0.17	url_length
-0.22	top10count_0
-0.38	top1count_0