

# PREDICTING QUERY DIFFICULTY ON THE WEB BY LEARNING VISUAL CLUES

Eric C. Jensen, Steven M. Beitzel,  
David Grossman, Ophir Frieder

Information Retrieval Laboratory  
Illinois Institute of Technology  
{ej,steve,grossman,frieder}@ir.iit.edu

Abdur Chowdhury

Search & Navigation Group  
America Online, Inc.  
cabdur@aol.com

## MOTIVATION

- Identifying queries that make it difficult for search engines to find relevant documents enables alternative processing strategies
- Query difficulty has been predicted for ad-hoc tasks based on ambiguity using clarity metrics
- How can we predict query difficulty in a precision-oriented web search task?

## METHODOLOGY

- Use visual features from retrieved surrogate document representations (titles, snippets, etc.)
- Perform supervised regression over counts of query terms in surrogate to predict mean score
- Does not require term or collection statistics

## EXPERIMENTATION

- Sampled 896 queries from AOL web search log
- Engines: Google, Yahoo, AltaVista, AllTheWeb, MSN, Lycos, Teoma, MSN Tech Preview, Wisenut, and Gigablast (anonymized here)
- Pooled all of the top 10 results from each
- AOL assessors and students judged binary relevance for every result using interface that displayed title and snippet and allowed clickthroughs, but hid the originating engine
- Collection available at <http://ir.iit.edu/collections>

## CONCLUSIONS

- Predictors from recall-oriented ad-hoc search (clarity) not as effective in precision-oriented tasks
- Visual features are much more effective at predicting relevance in a high precision task than query-based features
- Regression outperformed binary relevance prediction for each individual result

## PREDICTORS AND MEAN P@10 CORRELATIONS FOR ALL ENGINES

predictor	over all 896 no learning	over test 300 using SMO reg
$\sigma_{idf}$	-0.0544	0.2458
$\frac{idf_{max}}{idf_{min}}$	-0.1788	0.2985
<i>Simp. Clarity Score</i>	0.2458	0.2625
<i>qDifficulty</i>	NA	0.2457
<b><i>vDifficulty</i></b>	NA	<b>0.5735</b>
<i>vDifficulty and qDifficulty</i>	NA	0.5633

## LEARNING USING vDIFFICULTY AND qDIFFICULTY COMBINED FEATURES FOR EACH ENGINE

engine	correlation over test 300 using SMOreg	avg % n-grams in title weight	best query-based weight
E1	0.5049	0.3805	0.1690
E2	0.4928	0.5074	-0.3664
E3	0.5413	0.5234	-0.2885
E4	0.5641	0.4157	-0.4290
E5	0.5072	0.4220	-0.3886
E6	0.5599	0.4484	0.1367
E7	0.5303	0.6160	-0.3086
E8	0.5399	0.6824	-0.1212
E9	0.5722	0.5909	-0.1401
E10	0.5055	0.7707	-0.2714

