

PROBABILISTIC FIELD MAPPING FOR PRODUCT SEARCH

Aman Berhane Ghirmatsion and Krisztian Balog
University of Stavanger

TASK: Product search task of the LL4IR Lab

- Given a keyword query, return a ranked list of products from a product catalog
- Products are described by a number of attributes (name, brand, categories, characters, etc.)
- Historical click and query information is available

GOALS

- Establish a probabilistic mapping from query terms to document terms
- Exploit historical click information

RETRIEVAL FRAMEWORK

- Probabilistic Retrieval Model for Semistructured Data (PRMS) [1]

Document language model estimation

$$P(t|\theta_d) = \sum_{f \in F} P(f|t)P(t|\theta_{d_f})$$

Estimating mapping probabilities

$$P(f|t) = \frac{P(t|f)P(f)}{P(t)} = \frac{P(t|f)P(f)}{\sum_{f' \in F} P(t|f')P(f')}$$

METHODS FOR ESTIMATING MAPPING PROBABILITIES

- Three specific instantiations of the PRMS model

	$P(f)$	$P(t f)$
Method 1	$\propto \text{NDCG}_f$	$\frac{1}{ F }$
Method 2	$\frac{1}{ F }$	$P_{ML}(t C_f)$
Method 3	$\propto \text{NDCG}_f$	$P_{ML}(t C_f)$

- **P(f)** Field priors can be uniform or set proportional to individual field performance (NDCG), based on historical CTR
- **P(t|f)** The probability of a term occurring in a given field can be uniform or based on term statistics (as in the original PRMS)

RESULTS

	Round #1					Round #2				
	Outcome	#l	%W	%L	%T	Outcome	#l	%W	%L	%T
Method 1	0.2827	699	7.7	19.6	72.7	0.4118	731	11.5	16.4	72.1
Method 2	0.3413	725	9.8	18.9	71.3	0.4389	757	10.4	13.3	76.2
Method 3	0.3277	665	8.7	17.9	73.4	0.4795	767	10.7	11.6	77.7

REFERENCES

[1] J. Kim, X. Xue, and W. B. Croft. A probabilistic retrieval model for semistructured data. In *Proceedings of ECIR 2009*.