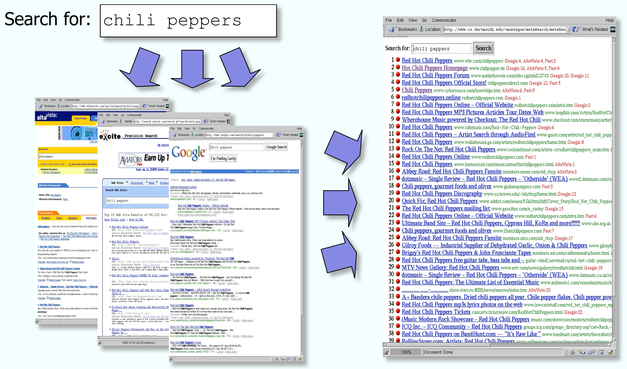
Measure-based Metasearch

Javed A. Aslam , Virgiliu Pavlu and Emine Yilmaz

Northeastern University

Introduction

The Metasearch Problem



Some Popular Metasearch Techniques

➢ **CombMNZ** is based on combining the normalized scores given to each document by the underlying systems.

➢**Condorcet** is based on viewing the metasearch problem as a multi-candidate election where the documents are candidates and the systems are voters expressing preferential rankings among the candidates.

Evaluation Measures

➢ **Average precision (AP)** is the average of the precisions at relevant documents.

➢ **Precision at cutoff *k* (PC(*k*))** is the precision at rank *k*.

➢ **R- precision (RP)** is the precision at rank *R*, where *R* is the number of documents relevant to the query.

Evaluation measures implicitly assign ***weights*** to the relevances of documents at various ranks, e.g. PC(10) implicitly assigns a weight of 1/10 to the relevances of the top 10 documents in a list and a weight of 0 to the remaining documents.

FACTS:

➢ Evaluation measures aim to assess how well a system retrieves relevant documents, as measured and evaluated by **implicit weights**.

➢Retrieval systems aim to retrieve relevant documents as well as possible.

HYPOTHESIS:

➢ Evaluation measures will assign **"high"** weights to relevant documents when applied to the lists generated by **"good"** retrieval systems. Thus,the weights implicit in evaluation measures can be used to identify likely relevant documents in a list.

Methodology

Weights Implicit in Evaluation Measures

➢Precisions at standard cutoffs (PC(*k*))

PC(*k*) implicitly assigns a weight of **1/*k*** to each of the top ***k*** documents in a list and a weight of **0** to every **remaining** document.

➢R-precision (RP)

R-precision is PC(*R*), where *R* is the total number of relevant documents for the query.

Problem : Computing the weights associated with R-precision requires *a priori* knowledge of *R*!

➢Average precision (AP)

AP is the average of the precisions at relevant documents. Hence, one can compute average precision as follows, where *N* is the length of the retrieved list, *rel(i)* is the 0-1 relevance of the document at rank *i*, *R* is the number of relevant documents in the query.

$$\begin{aligned} \text{AP} &= \frac{1}{R} \cdot \sum_{i\,:\,rel(i)=1} \text{PC}(i) \\ &= \frac{1}{R} \cdot \sum_{i=1}^{N} rel(i) \cdot \text{PC}(i) \\ &= \frac{1}{R} \cdot \sum_{i=1}^{N} rel(i) \sum_{j=1}^{i} rel(j)/i \\ &= \frac{1}{R} \cdot \sum_{1 \le j \le i \le N} \frac{1}{i} \cdot rel(i) \cdot rel(j) \end{aligned}$$

Thus, average precision assigns a weight of $\frac{1}{R \cdot i}$ to each pair of ranks (i, j), for all $1 \le j \le i \le N$.

We can compute the ***weight*** associated with each rank ***r*** by summing the weights associated with ***all pairs involving r***:

$$\begin{aligned} \sum_{j=1}^{r} \frac{1}{R \cdot r} + \sum_{i=r+1}^{N} \frac{1}{R \cdot i} &= \frac{1}{R} \cdot \left(1 + \frac{1}{r+1} + \frac{1}{r+2} + \cdots + \frac{1}{N}\right) \\ &= \frac{1}{R} \cdot (1 + H_N - H_r) \end{aligned}$$

where H_k is the *k*-th harmonic number.

Note: *R simply acts as a scaling factor. We don't need* a priori *knowledge of R*!

Metasearch Lists Through Implicit Weights

To obtain a consensus metasearch list through the weights assigned by an evaluation measure:

➢ compute the average weight assigned to the document across the underlying lists.

➢ rank the documents according to these average scores.

Results

TREC	MNZ	COND	AP	RP	PC(10)	PC(20)	PC(100)	PC(1000)
5	.294	.307	.300$^{\bullet}$	.308	.265	.280	.294	.237
6	.341	.315	.344°	.357$^{\bullet\circ}$	.292	.315	.341	.312
7	.320	.308	.333$^{\bullet\circ}$	.334°	.295	.311	.331	.301
8	.350	.343	.370$^{\bullet\circ}$	.366$^{\bullet\circ}$	.286	.313	.357	.305
9	.351	.348	.345	.353	.309	.303	.323	.259

The table shows the mean average precisions of the metasearch algorithms associated with average precision, R-precision and precision at cutoff *k*, where *k* =10, 20, 100, 1000 and compares those with CombMNZ And Condorcet algorithms. The results show that:

➢System-oriented measures, e.g. AP and RP, tend to implicitly weight the set of relevant documents higher than user-oriented measures, e.g.PC(*k*).

➢Performance of the metasearch algorithms corresponding to AP and RP often equals or exceeds CombMNZ and Condorcet.

Note: *The AP and RP results shown in the table constitute statistically significant improvements*[1] *w.r.t. CombMNZ and Condorcet when labeled with a* $\bullet$ *and/or* $\circ$.

[1]Sign test of significance, 90% confidence level.