



**INLS 490-154: Information Retrieval Systems
Design & Implementation
Spring 2009**

Evaluation-I

Art of identifying good, bad and ugly

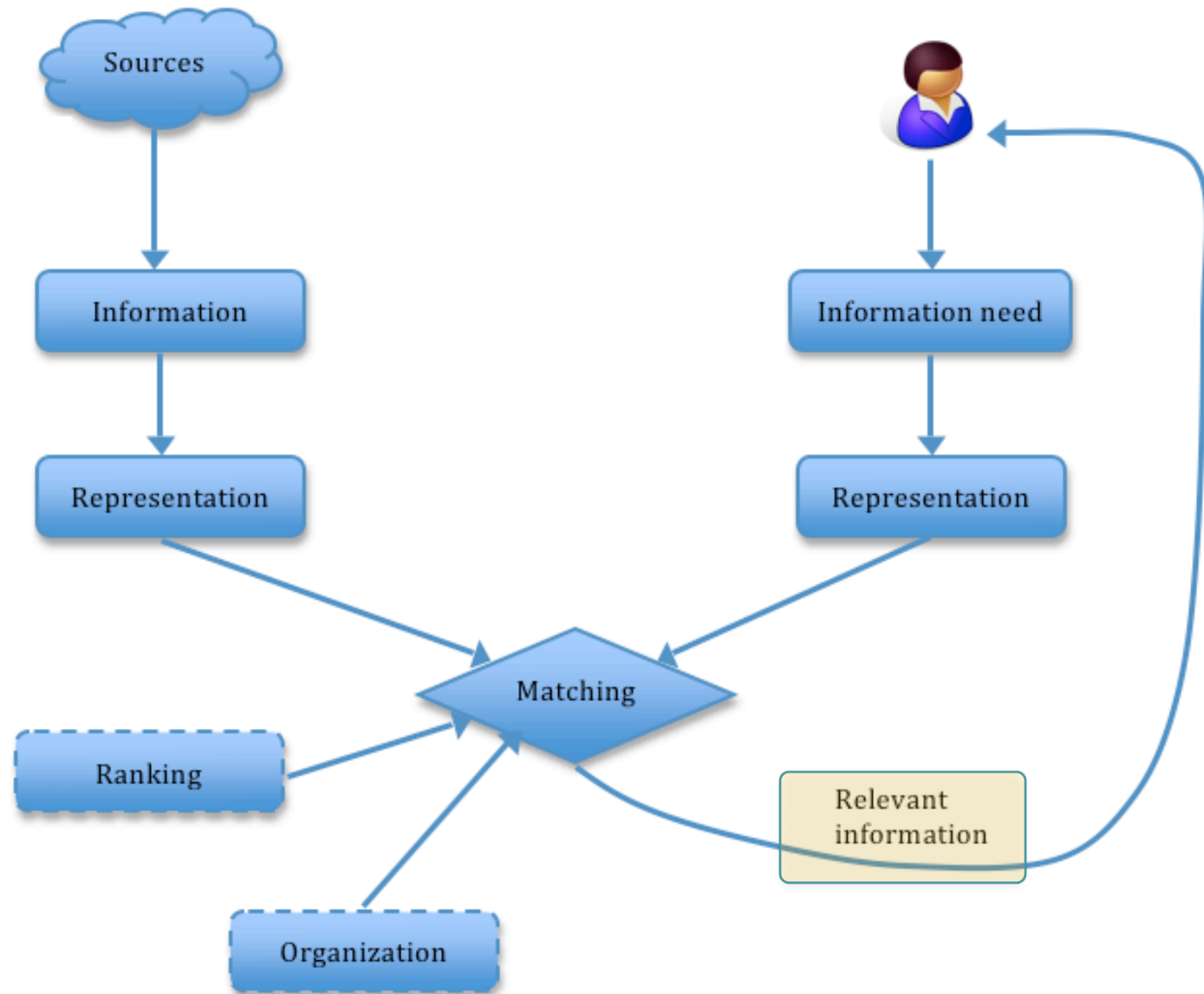
February 26, 2009

Chirag Shah

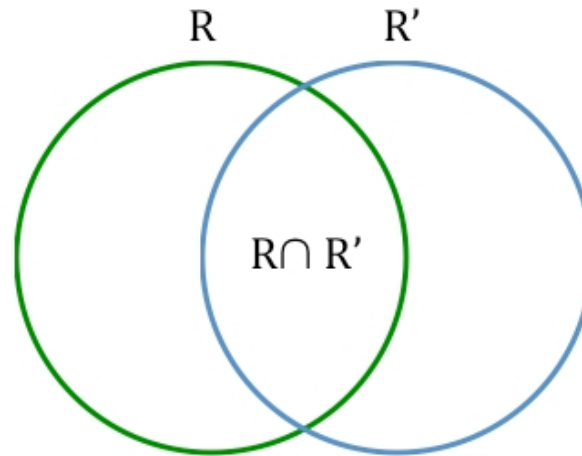
School of Information & Library Science (SILS)

UNC Chapel Hill

Today's lab



Recall and precision



R= Relevant, R'=Retrieved

$$\text{Recall} = \frac{R \cap R'}{R}$$

$$\text{Precision} = \frac{R \cap R'}{R'}$$

Recall and precision extended to rank list

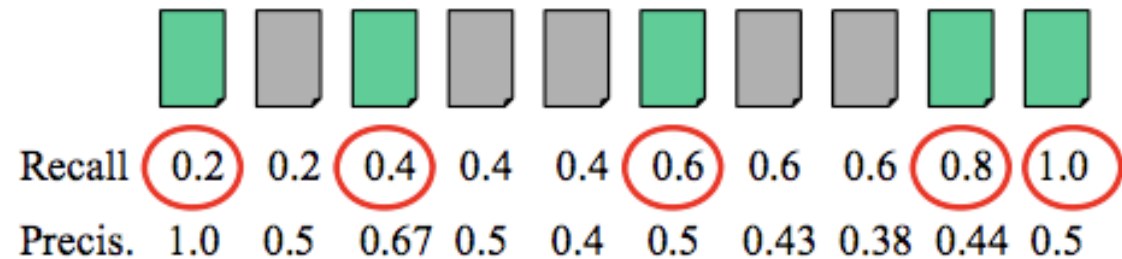
- Create sets going down the rank list
 - At every document
 - At an “interesting” point
 - At every relevant document
 - At fixed rank value cutoff
 - E.g. precision at rank 10
 - At fixed recall points
 - E.g. precision at 20% recall

Recall and precision examples

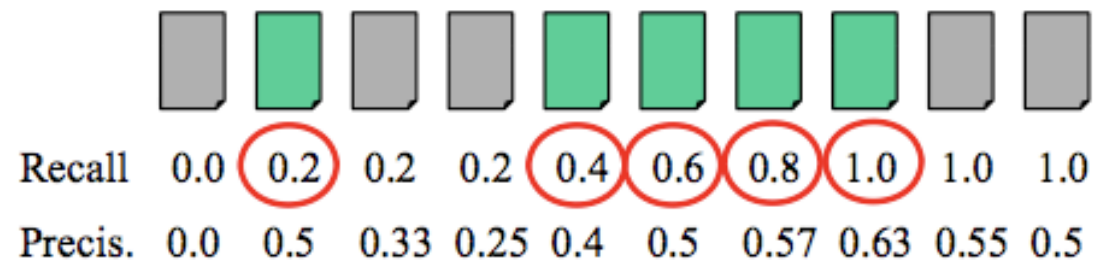


= the relevant documents

Ranking #1



Ranking #2



Source: James Allan, UMass Amherst

Average precision

- Calculate by averaging precision when recall increases.

Recall	0.2	0.2	0.4	0.4	0.4	0.6	0.6	0.6	0.8	1.0	
Precis.	1.0	0.5	0.67	0.5	0.4	0.5	0.43	0.38	0.44	0.5	AvgPrec= 62.2%

Recall	0.0	0.2	0.2	0.2	0.4	0.6	0.8	1.0	1.0	1.0	
Precis.	0.0	0.5	0.33	0.25	0.4	0.5	0.57	0.63	0.55	0.5	AvgPrec= 52.0%

- Mean Average Precision (MAP) = average of average precision over all the queries

Source: James Allan, UMass Amherst

R-precision

- Precision after R documents retrieved, where R is the total number of relevant documents for a given query.
- Gives one number to compare.
- Average precision and R-precision are shown to be highly correlated.



Evaluation exercise

1. Obtain LA Times data from TREC collection
2. Index the data (BuildIndex)
3. Obtain topics
4. Prepare query file
5. Parse the query file (ParseToFile)
6. Retrieval (RetEval)
7. Evaluate

Evaluating using trec_eval

trec_eval <options> <rel_file> <ret_file>

- **-o**: Print requested measures in old non-relational format
- **-q**: In addition to summary evaluation, give evaluation for each query
- **-a**: Print all evaluation measures, instead of just official measures



Summary

- Recall and precision are the most popular measure of evaluation in IR.
- On average, as recall increases, precision drops, and vice versa.
- Average precision, mean average precision (MAP), and R-precision provide us with single numbers to compare retrieval performance.



Next time

- Continue exploring recall and precision measures.
- Learn about other evaluation metrics.
- Compare retrieval performances.