Discussion Session
Week 9

INF 141: Information Retrieval
Winter 2009

Yasser Ganjisaffar

yganjisa@ics.uci.edu

Outline

- Assignment 06 Questions
- Web Search Evaluation

Failures!

- Task attempt_200903011033_0154_m_000185_0 failed to report status for 600 seconds. Killing!

- org.apache.hadoop.util.DiskChecker$DiskErrorException: Could not find any valid local directory for taskTracker/jobcache/job_200903011033_0154/jars
  - Machine: http://carter-pewterschmidt.ics.uci.edu:50060/

- java.lang.NoClassDefFoundError: edu/uci/ics/crawler4j/crawler/HTMLParser
  - Machine: http://carter-pewterschmidt.ics.uci.edu:50060/

- KILLED

How to Evaluate Search Results?

● … … … … … … … … … …

● … … … … … … … … … …

● … … … … … … … … … …

● … … … … … … … … … …

● … … … … … … … … … …

● … … … … … … … … … …

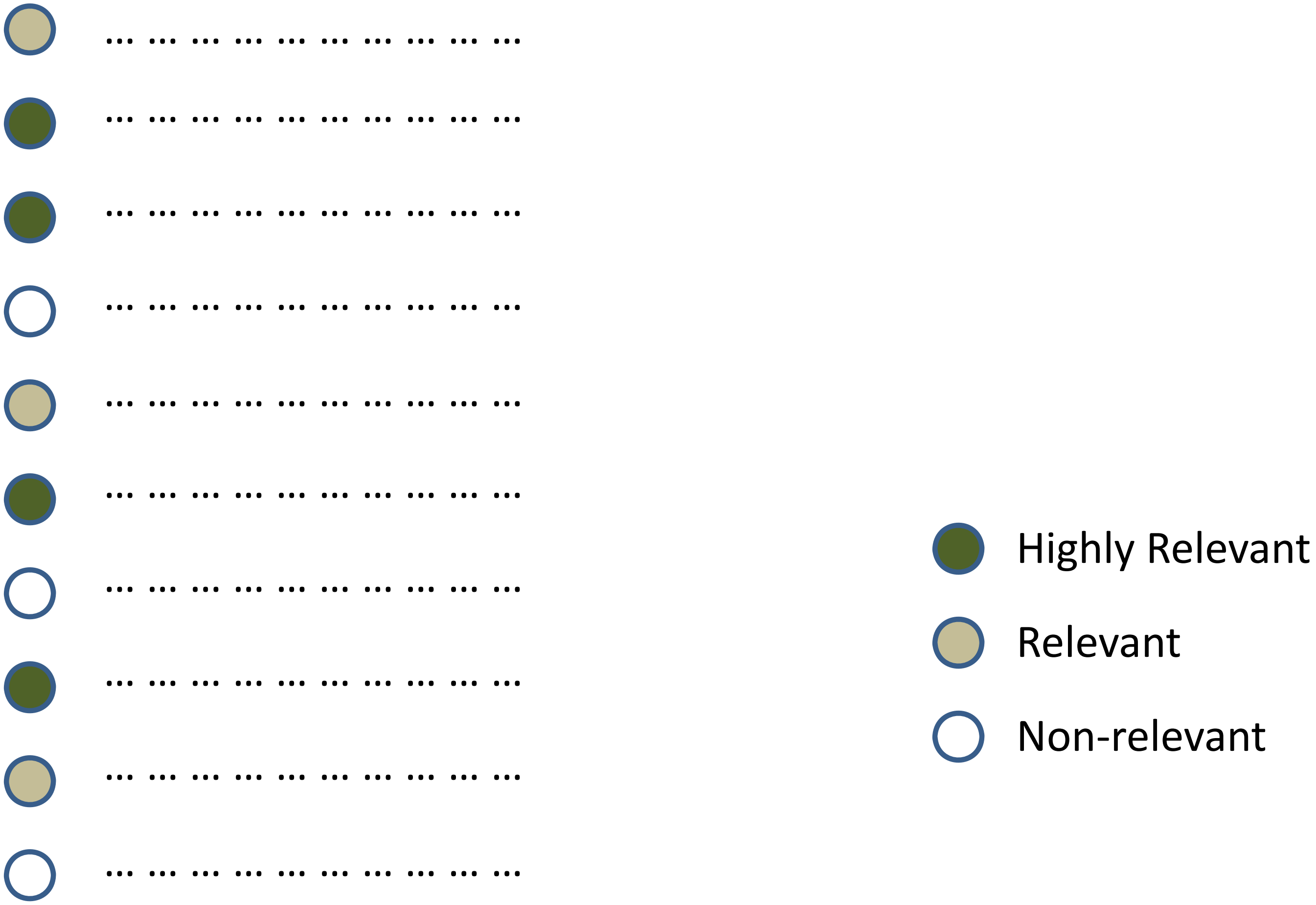
● … … … … … … … … … …

● … … … … … … … … … …

● … … … … … … … … … …

● … … … … … … … … … …

Expert Labeling of Search Results

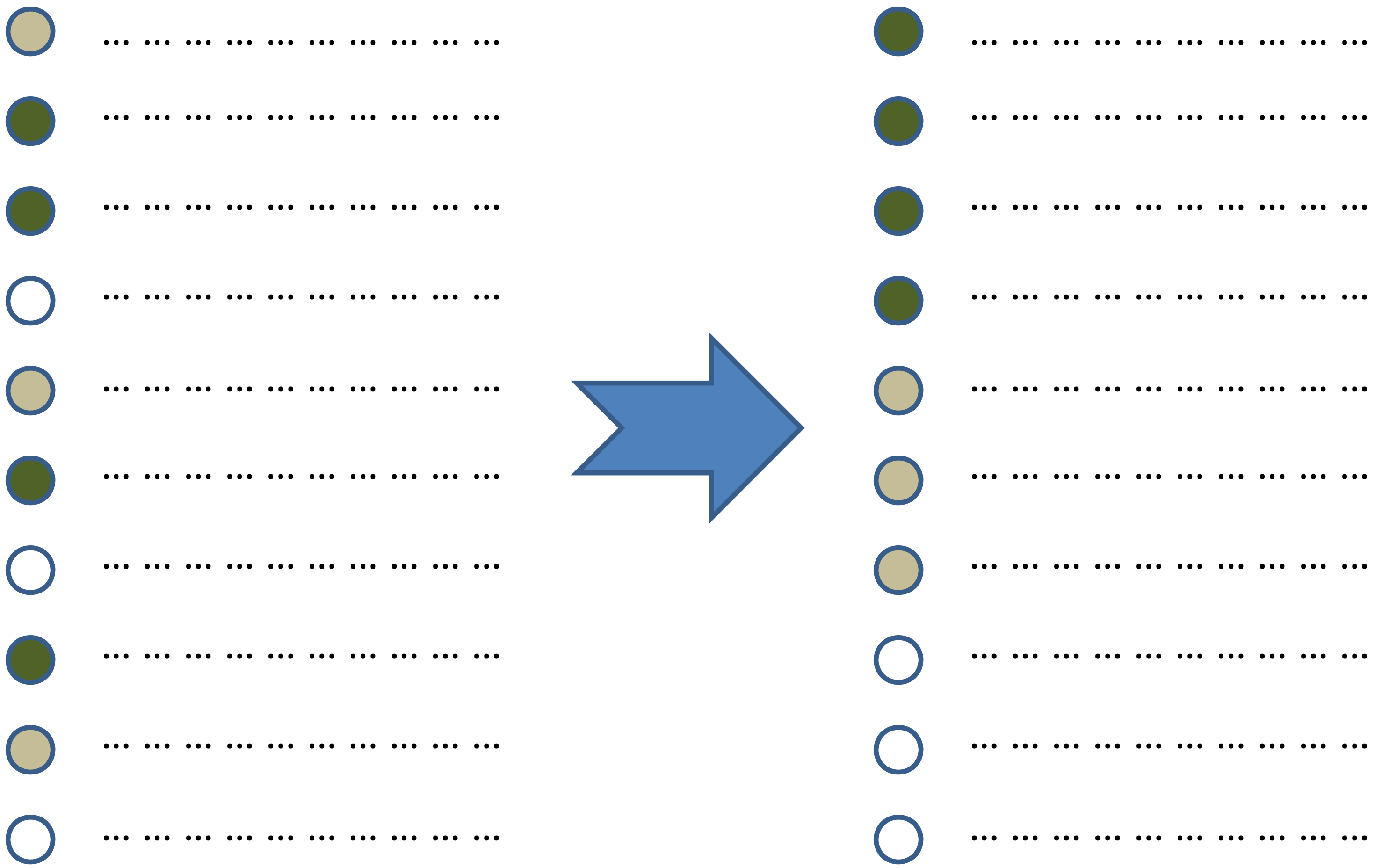


● Highly Relevant

◍ Relevant

○ Non-relevant

Ideal Ranking of Results

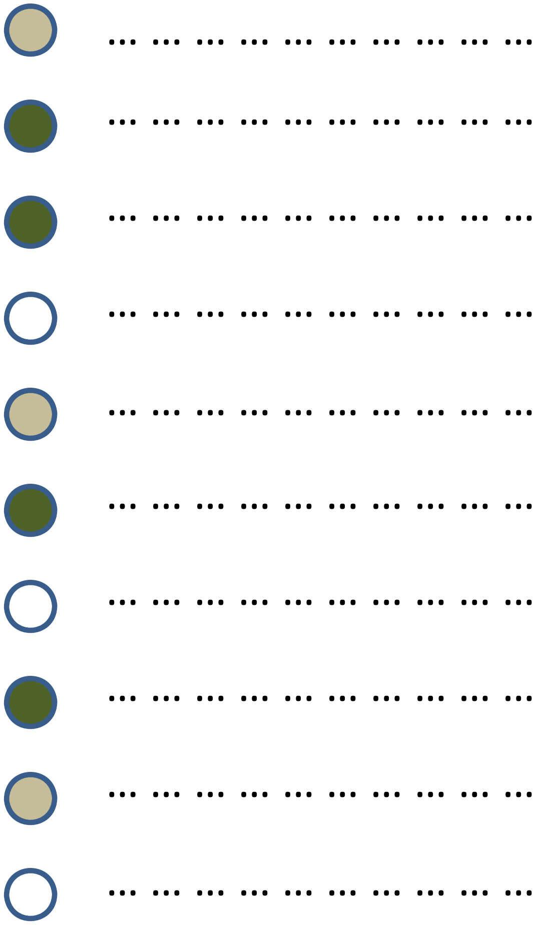


Search Results

Ideal Ranking

How to Compare *Current Ranking* with *Ideal Ranking*?

Cumulative Gain (CG)



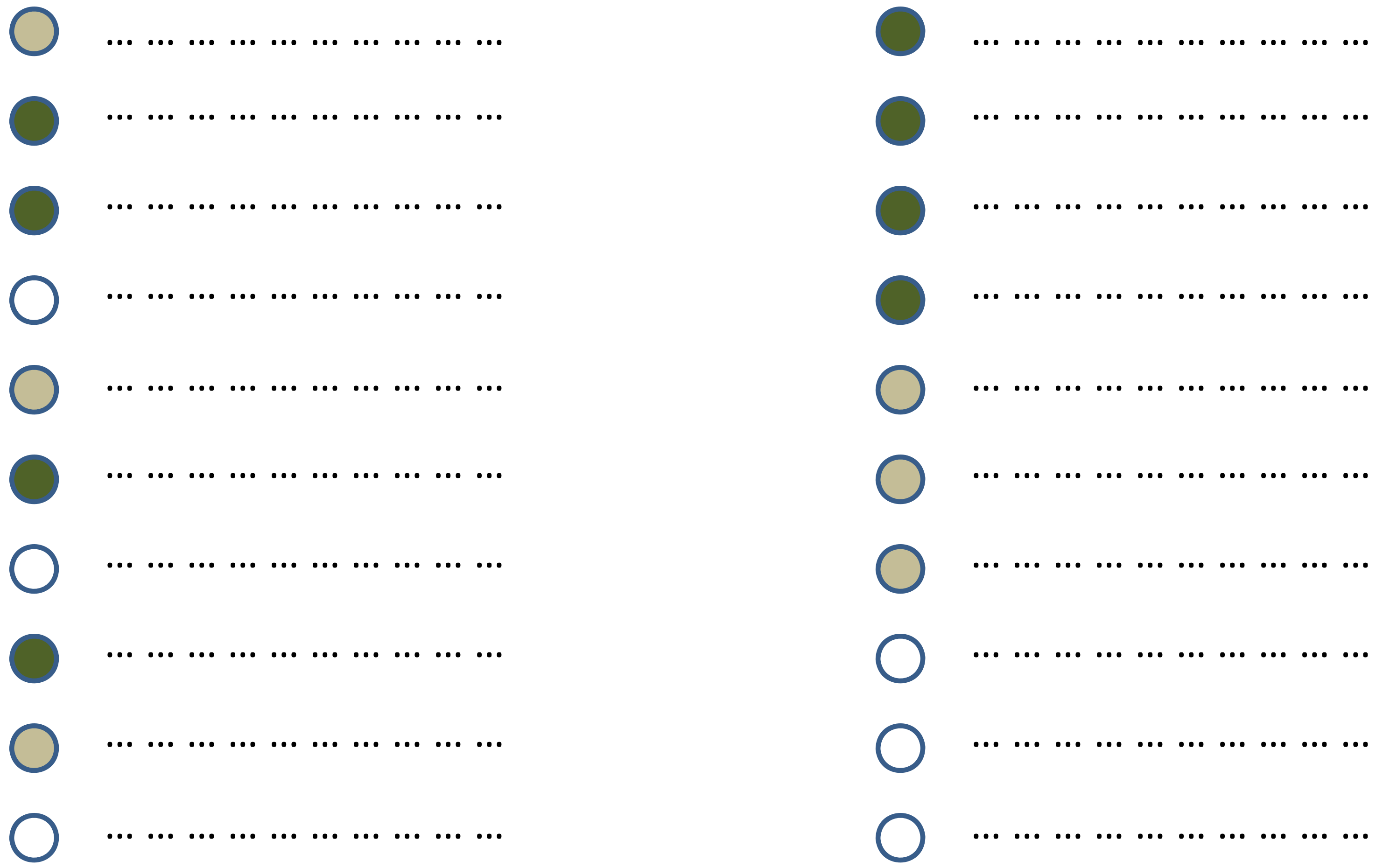
Type	Gain
Highly Relevant	2
Relevant	1
Non-relevant	0

$CG_{10} = 4 \text{ x } 2 + 3 \text{ x } 1 + 3 \text{ x } 0 = 11$

$CG_{5} = 2 \text{ x } 2 + 1 \text{ x } 1 + 1 \text{ x } 0 = 5$

$CG_{2} = 1 \text{ x } 2 + 1 \text{ x } 1 = 3$

Cumulative Gain



$CG_{10} = 4 \times 2 + 3 \times 1 + 3 \times 0 = 11$

$CG_{10} = 4 \times 2 + 3 \times 1 + 3 \times 0 = 11$

Discounted Cumulative Gain

- Assumptions:
  - Highly relevant documents are more useful when appearing earlier in a search engine result list (have higher ranks).
  - Highly relevant documents are more useful than marginally relevant documents, which are in turn more useful than irrelevant documents.
- Measures the *gain* of a document based on its position in the result list.

Discounted Cumulative Gain

$$\mathrm{DCG_p} = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{log_2 i}$$

$$\mathrm{DCG_p} = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{log_2(1 + i)}$$

DCG Calculations

◉ … … … … … … … … … … 1/log(2) = 1

● … … … … … … … … … … 3/log(3) = 1.9

● … … … … … … … … … … 3/log(4) = 1.5

○ … … … … … … … … … … 0

◉ … … … … … … … … … … 1/log(6) = 0.4

● … … … … … … … … … … 3/log(7) = 1.1

○ … … … … … … … … … … 0

● … … … … … … … … … … 3/log(9)= 0.9

◉ … … … … … … … … … … 1/log(10)= 0.3

○ … … … … … … … … … … 0

$$\frac{2^{r(p)} - 1}{\log(1 + p)}$$

DCG$_{10}$ = 7.1

Ideal DCG

● … … … … … … … … … … 3/log(2) = 3

● … … … … … … … … … … 3/log(3) = 1.9

● … … … … … … … … … … 3/log(4) = 1.5

● … … … … … … … … … … 3/log(5) = 1.3

◍ … … … … … … … … … … 1/log(6) = 0.4

◍ … … … … … … … … … … 1/log(7) = 0.36

◍ … … … … … … … … … … 1/log(8) = 0.33

○ … … … … … … … … … … 0

○ … … … … … … … … … … 0

○ … … … … … … … … … … 0

$$\frac{2^{r(p)} - 1}{\log(1 + p)}$$

IDCG_{10} = 8.79

Normalized DCG

◍ … … … … … … … … … … 1/log(2) = 1

● … … … … … … … … … … 3/log(3) = 1.9

● … … … … … … … … … … 3/log(4) = 1.5

○ … … … … … … … … … … 0

◍ … … … … … … … … … … 1/log(6) = 0.4

● … … … … … … … … … … 3/log(7) = 1.1

○ … … … … … … … … … … 0

● … … … … … … … … … … 3/log(9)= 0.9

◍ … … … … … … … … … … 1/log(10)= 0.3

○ … … … … … … … … … … 0

$$\mathrm{nDCG_p} = \frac{DCG_p}{IDCGp}$$

nDCG_{10} = 7.1/8.79 = 81%

Drawback of DCG?

- Labeling results is expensive.
- No ideal ordering of results when only partial relevance feedback (labels) is available.

Click-through Data: Implicit Feedback

1 ○ … … … … … … … … … …

2 ● … … … … … … … … … …

3 ○ … … … … … … … … … …

4 ○ … … … … … … … … … …

5 ● … … … … … … … … … …

6 ○ … … … … … … … … … …

7 ● … … … … … … … … … …

8 ○ … … … … … … … … … …

9 ○ … … … … … … … … … …

10 ○ … … … … … … … … … …

Assuming that user has checked results from top to bottom:

2 is more relevant than 1.

5 is more relevant than 1, 3, 4

7 is more relevant than 1, 3, 4, 6

(2,1) (5,1) (5,3) (5,4)

(7,1) (7,3) (7,4) (7,6)

Learning to Rank

(2,1)

(5,1)

(5,3)

(5,4)

(7,1)

⋮

- An ideal search engine should rank “2” higher than “1”.

- We can use this training data to learn how to rank search results.

Learning to Rank

	TF-IDF1	TF-IDF2	PageRank1	PageRank2	Age1	Age2	Title Score1	Title Score2	…
(2,1)									
(5,1)									
…									

Google uses more than 200 features