

Queries (cont.)

Query suggestions

Query Suggestions

- Reformulation suggestions
- Autocomplete
- Google Instant
 - How would you make Google Instant?
 - How does this affect user reformulations?

Exercise and break

- For some queries, many users are not clicking on any results.
- Larry says this is because they did not see any relevant documents in the results.
- Sergey says this is because they found the answers on the results page.
- To be continued... Nov 18

User vs Search Engines

- User reformulations as feedback
- What can the search engine do when it notices users reformulating?
- Users learn from returned results, search engines learn from user behavior

A Quick Primer on Information Theory

Information Entropy

Why does English text compress better than an image?

- Information entropy
 - Expected value of the information in a message (file)
- Is a picture worth a thousand words?
 - Example

Relative Entropy

- Also known as information gain
- Also known as kl-divergence
 - Example
 - 2 “catches”

Evaluation

Was it good for you?

How do we evaluate a
search engine?

List every component/aspect of the search system you think that it can be evaluated:

(the list below is the one that was generated by another class)

- Response time
- Relevance of results
- Accuracy of results
- Number of results
- Ease of use of UI
- Credibility of results
- Currency of results
- Form of returned results
 - Video, text, mp3, ...
- Organization of results
 - Ranking
 - Grouping or A-Z order
- Filtering of results
 - Narrowing by field
- Clarity of results page
 - Snippets & other info that describes the results
- Query
 - Compare to information need
 - Specificity
 - Relation to results
 - Query logs
- Type of query syntax
 - Boolean
 - Advanced operators
- Query processor
 - Transformations
- Content
 - Type
 - Quality
- Inverted Index size & quality
- Matching
 - Overall Performance
 - At different fields
 - Efficiency

How do we evaluate search engine **results**?

- Relevance of results
- Diversity of results
 - Entropy

Relevance of Results

1. Humans label each pair: (query, document)
 - Which queries to label?
 - Head vs tail queries
 - How do we know if their labels are good?
 - Check agreement using Kappa values
2. Use some evaluation metric
 - Precision and recall

Calculating Precision

- How many of the documents retrieved are relevant?

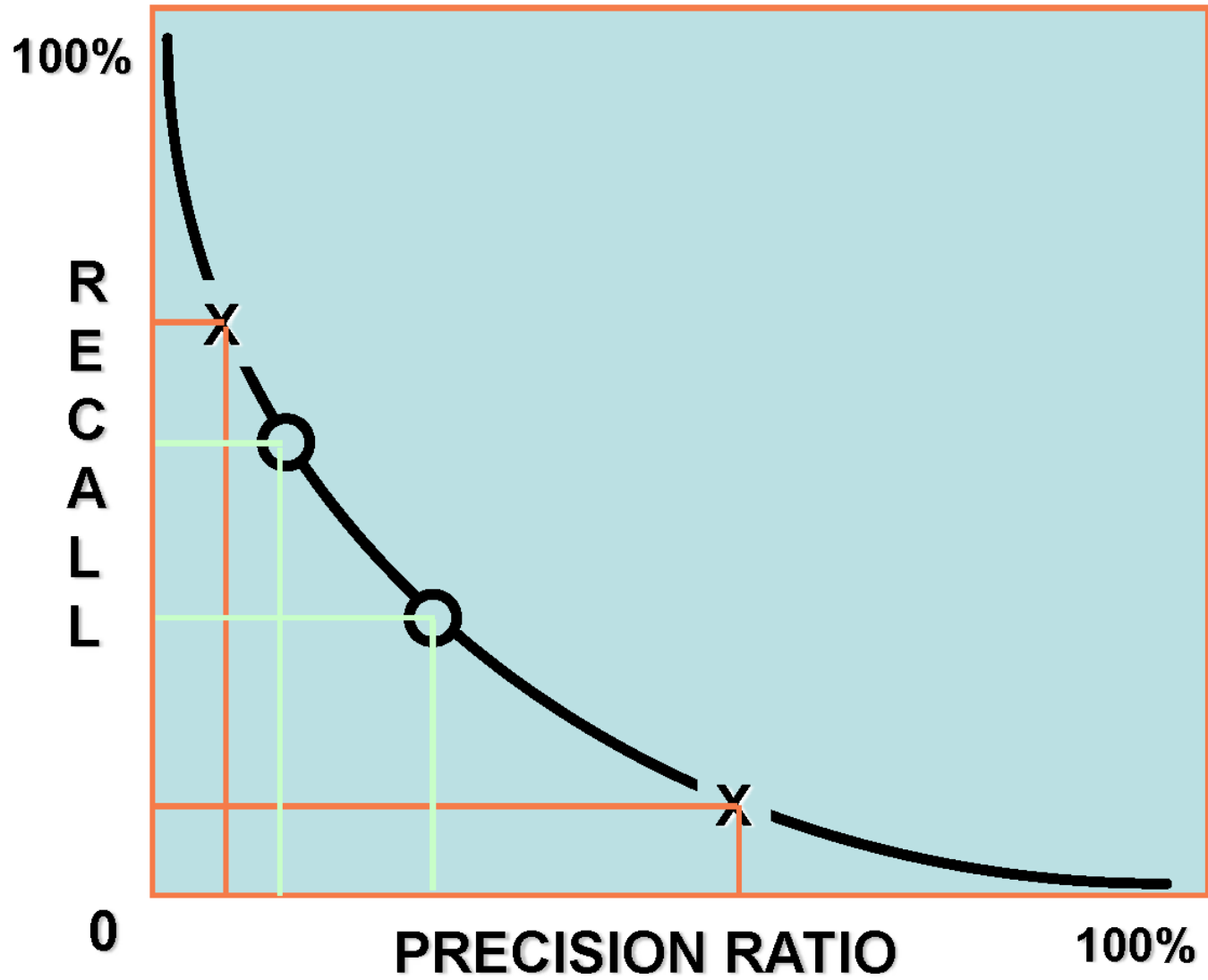
$$\text{precision} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{retrieved documents}\}|}$$

Calculating Recall

- Out of all the relevant documents, how many did the search engine retrieve?

$$\text{recall} = \frac{|\{\text{relevant documents}\} \cap \{\text{retrieved documents}\}|}{|\{\text{relevant documents}\}|}$$

R-P Relationship



Calculating F-measure

- Balance between precision and recall
- For perfect balance,

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}$$

More Evaluation

- What if we want to use relevance and rank in the evaluation metric??

Evaluation Metrics

| Scoring? | Retrieval Technique | Evaluation Technique |
|----------|---------------------|----------------------|
| No | Boolean | Precision and Recall |
| Yes | Ranked Results | MAP, nDCG |

MAP (Mean Average Precision)

- Precision adjusted by rank

$$\text{AveP} = \frac{\sum_{r=1}^N (P(r) \times \text{rel}(r))}{\text{number of relevant documents}}$$

$$P(r) = \frac{|\{\text{relevant retrieved documents of rank } r \text{ or less}\}|}{r}$$

Cumulative Gain

- Sum of all relevance scores of returned results (up to p)

$$CG_p = \sum_{i=1}^p rel_i$$

Discounted Cumulative Gain

- Like CG, but takes rank into account

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

nDCG

- Normalized discounted cumulative gain
- Like DCG, but normalizes for each query

$$\text{nDCG}_p = \frac{DCG_p}{IDCG_p}$$