

Web Search: Techniques, algorithms and Applications

Basic Techniques for Web Search

German Rigau <german.rigau@ehu.es>

[Based on slides by Eneko Agirre ...
and Christopher Manning and Prabhakar Raghavan]



Basic Techniques for Web Search

- Review of applications
- Basic Techniques in detail:
 - Boolean search
 - Vocabularies, dictionaries, index
 - **Scoring, evaluation, complete system**
 - Web search
- Semantic search

Evaluating Search Engines

- ...

Measures for a search engine

- How fast does it index
 - Number of documents/hour
 - (Average document size)
- How fast does it search
 - Latency as a function of index size
- Expressiveness of query language
 - Ability to express complex information needs
 - Speed on complex queries
- Uncluttered UI
- Is it free?

Measures for a search engine

- All of the preceding criteria are *measurable*: we can quantify speed/size
 - we can make expressiveness precise
- The key measure: user happiness
 - What is this?
 - Speed of response/size of index are factors
 - But blindingly fast, useless answers won't make a user happy
- Need a way of quantifying user happiness

Measuring user happiness

- Issue: who is the user we are trying to make happy?
 - Depends on the setting
- Web engine:
 - User finds what they want and return to the engine
 - Can measure rate of return users
 - User completes their task – search as a means, not end
 - See Russell
<http://dmrussell.googlepages.com/JCDL-talk-June-2007-short.pdf>
- eCommerce site: user finds what they want and buy
 - Is it the end-user, or the eCommerce site, whose happiness we measure?
 - Measure time to purchase, or fraction of searchers who become buyers?

Measuring user happiness

- Enterprise (company/govt/academic): Care about “user productivity”
 - How much time do my users save when looking for information?
 - Many other criteria having to do with breadth of access, secure access, etc.

Happiness: elusive to measure

- Most common proxy: *relevance* of search results
- But how do you measure relevance?
- We will detail a methodology here, then examine its issues
- Relevance measurement requires 3 elements:
 1. A benchmark document collection
 2. A benchmark suite of queries
 3. A usually binary assessment of either Relevant or Nonrelevant for each query and each document
 - Some work on more-than-binary, but not the standard

Evaluating an IR system

- Note: the **information need** is translated into a **query**
- Relevance is assessed relative to the **information need** *not* the **query**
- E.g., Information need: *I'm looking for information on whether drinking red wine is more effective at reducing your risk of heart attacks than white wine.*
- Query: **wine red white heart attack effective**
- You evaluate whether the doc addresses the information need, not whether it has these words

Standard relevance benchmarks

- TREC - National Institute of Standards and Technology (NIST) has run a large IR test bed for many years
- Reuters and other benchmark doc collections used
- “Retrieval tasks” specified
 - sometimes as queries
- Human experts mark, for each query and for each doc, Relevant or Nonrelevant
 - or at least for subset of docs that some system returned for that query

Unranked retrieval evaluation: Precision and Recall

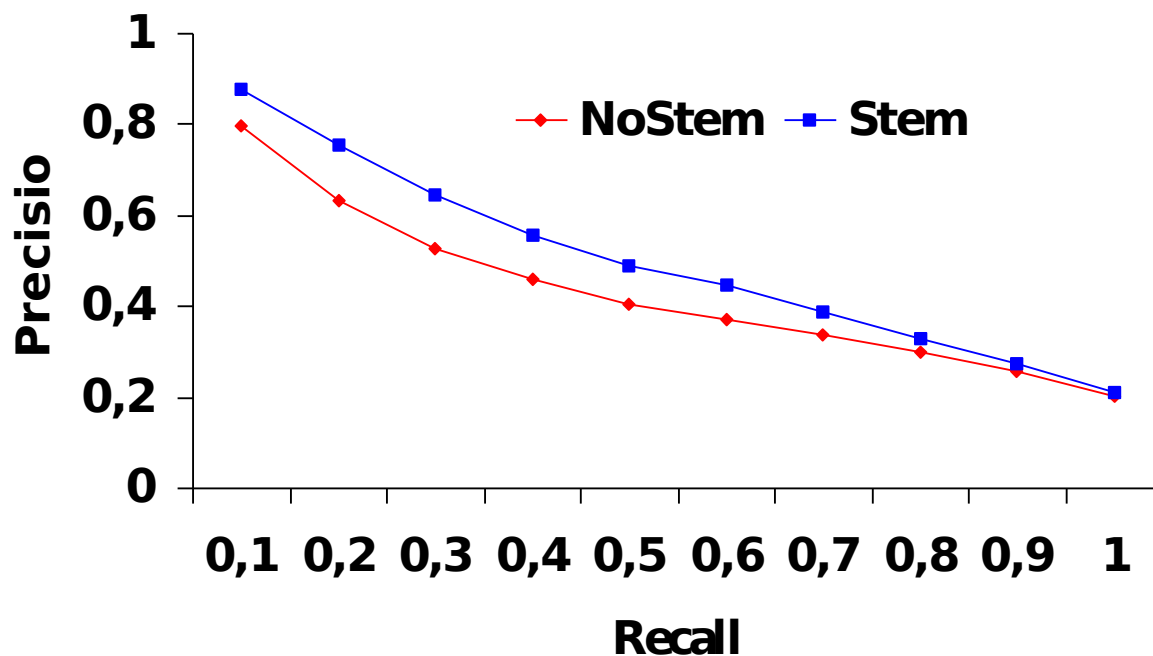
- **Precision:** fraction of retrieved docs that are relevant = $P(\text{relevant}|\text{retrieved})$
- **Recall:** fraction of relevant docs that are retrieved = $P(\text{retrieved}|\text{relevant})$

	Relevant	Nonrelevant
Retrieved	tp	fp
Not Retrieved	fn	tn

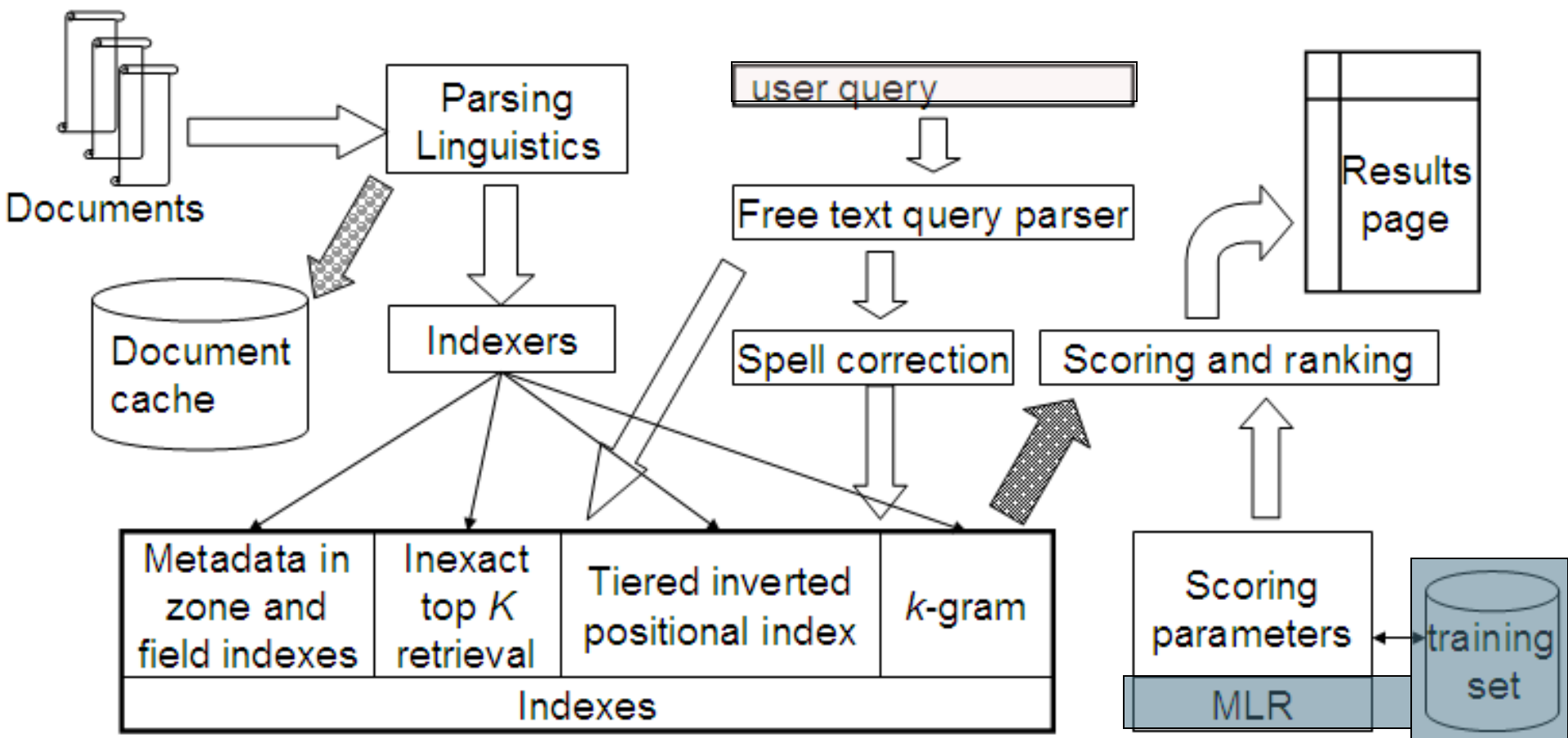
- Precision $P = tp/(tp + fp)$
- Recall $R = tp/(tp + fn)$

How To Compare Two or More Systems

- The curve closest to the upper right-hand corner of the graph indicates the best performance



Putting it all together



Web Search: Techniques, algorithms and Applications

Basic Techniques for Web Search

German Rigau <german.rigau@ehu.es>

[Based on slides by Eneko Agirre ...
and Christopher Manning and Prabhakar Raghavan]

