Modeling User Behavior and Interactions

Lecture 5: Search Interfaces + New Directions

Eugene Agichtein
Emory University
Lecture 5 Plan

1. Generating result summaries (abstracts)
   – Beyond result list

2. Spelling correction and query suggestion

3. New directions in search user interfaces
   – Collaborative Search
   – Collaborative Question Answering

• PhD studies in the U.S. (and in Emory U)
1. Generating Result Summaries

• How to present search results list to a user?

• Most commonly, a list of the document titles plus a short summary, aka “10 blue links”
Good Summary Guidelines

• All query terms should appear in the summary, showing their relationship to the retrieved page
• When query terms are present in the title, they need not be repeated
  – allows snippets that do not contain query terms
• Highlight query terms in URLs
• Snippets should be readable text, not lists of keywords
How to Generate Good Summaries?

• The title is typically automatically extracted from document metadata. What about the summaries?
  – This description is crucial.
  – User can identify good/relevant hits based on description.

• Two main kinds of summaries:
  – **Static summary**: always the same, regardless of the query that hit the doc
  – **Dynamic summary**: *query-dependent* attempt to explain why the document was retrieved for the query at hand
Dynamic Summary Generation

**Tropical Fish**
One of the U.K.'s Leading suppliers of Tropical, Coldwater, Marine Fish and Invertebrates plus... next day fish delivery service...

www.tropicalfish.org.uk/tropical_fish.htm  Cached page

- Query-dependent document summary
- Simple summarization approach
  - rank each sentence in a document using a significance factor
  - select the top sentences for the summary
  - first proposed by Luhn in 50's
Sentence Selection

• Significance factor for a sentence is calculated based on the occurrence of *significant words*
  
  – If $f_{d,w}$ is the frequency of word $w$ in document $d$, then $w$ is a significant word if it is not a stopword and

\[
    f_{d,w} \geq \begin{cases} 
    7 - 0.1 \times (25 - s_d), & \text{if } s_d < 25 \\
    7, & \text{if } 25 \leq s_d \leq 40 \\
    7 + 0.1 \times (s_d - 40), & \text{otherwise}
    \end{cases}
\]

  where $s_d$ is the number of sentences in document $d$

  – text is *bracketed* by significant words (limit on number of non-significant words in bracket)
Sentence Selection

• Significance factor for bracketed text spans is computed by dividing the square of the number of significant words in the span by the total number of words

• e.g.,

\[
\begin{align*}
\text{w w w w w w w w w w w w.} \\
\text{(Initial sentence)} \\
\text{w w s w s s w w s w w w.} \\
\text{(Identify significant words)} \\
\text{w w [s w s s w w s] w w.} \\
\text{(Text span bracketed by significant words)}
\end{align*}
\]

• Significance factor = \(4^2/7 = 2.3\)
Dynamic Snippet Generation (Cont’d)

• Involves more features than just significance factor
• e.g. for a news story, could use
  – whether the sentence is a heading
  – whether it is the first or second line of the document
  – the total number of query terms occurring in the sentence
  – the number of unique query terms in the sentence
  – the longest contiguous run of query words in the sentence
  – a density measure of query words (significance factor)

• Weighted combination of features used to rank sentences
Static Summary Generation

- Web pages are less structured than news stories
  - can be difficult to find good summary sentences
- Snippet sentences are often selected from other sources
  - metadata associated with the web page
    - e.g., <meta name="description" content= ...>
  - external sources such as web directories
    - e.g., Open Directory Project, http://www.dmoz.org
  - Wikipedia: summary paragraph, infoboxes, ...
Problem? Very Good Summaries May Not Get **Clicks**!

Everything you needed is in the summary

Eugene Agichtein, Emory University, RuSSIR 2009 (Petrozavodsk, Russia)
Organizing Search Results

Dumais, S., E. Cutrell, and H. Chen. Optimizing search by showing results in context, CHI 2001

List Organization Query: jaguar Category Org (SWISH)

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
System Components

Dumais, S, E. Cutrell, and H. Chen. *Optimizing search by showing results in context*, CHI 2001

- **Training (offline)**: classified web pages → SVM → model
- **Running (online)**: web search results → SVM → classified search results
Text Classification

- Dumais, S., E. Cutrell, and H. Chen. *Optimizing search by showing results in context*, CHI 2001

- **Text Classification**
  - Assign documents to one or more of a predefined set of categories
  - E.g., News feeds, Email - spam/no-spam, Web data
  - Manually vs. automatically

- **Inductive Learning for Classification**
  - Training set: Manually classified a set of documents
  - Learning: Learn classification models
  - Classification: Use the model to automatically classify new documents
• Support Vector Machine (SVM)
  – Accurate and efficient for text classification (Dumais et al., Joachims)
  – Model = weighted vector of words
    • “Automobile” = motorcycle, vehicle, parts, automobile, harley, car, auto, honda, porsche ...
    • “Computers & Internet” = rfc, software, provider, windows, user, users, pc, hosting, os, downloads ...

• Hierarchical Models
  – 1 model for N top level categories
  – N models for second level categories
  – Very useful in conjunction w/ user interaction
Information Overlay

**Dumais, S, E. Cutrell, and H. Chen.** *Optimizing search by showing results in context*, CHI 2001

- Use tooltips to show
  - Summaries of web pages
  - Category hierarchy

---

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Expansion of Category Structure

**Dumais**, S., E. Cutrell, and H. Chen. *Optimizing search by showing results in context*, CHI 2001

- **Automotive**
  - **Maintenance & Repair**
    - (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
    - (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
  - **Buy or Sell a Car**
    - (85) Bauer Jaguar, your specialist in luxury foreign sports cars and Jaguar automob
    - (84) A&L Luxury Car Center - Jaguar Main Page
  - **Vintage & Classic**
    - (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot Ferrari Fiat B
  - **Makes, Models & Clubs**
    - (94) Jaguar Club of Florida

Eugene Agichtein, Emory University, RuSSIR 2009 (Petrozavodsk, Russia)
User Study - Conditions

Dumais, S., E. Cutrell, and H. Chen. *Optimizing search by showing results in context*, CHI 2001

**Category Interface**

- Computers & Internet
  - SubCategory
  - More (18)
    - (92) Clan Jaguar Quake & Quake 2 Clan
    - (90) Team Jaguar-System
    - (76) Team Jaguar Quake Teams
    - (69) Jaguar 2000 Screen Saver

- Automotive
  - SubCategory
  - More (16)
    - (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot
      - Ferrari Fiat B
    - (94) Jaguar Club of Florida
    - (85) Eager Jaguar, your specialist in luxury foreign sports cars and Jaguar automobi
    - (94) A.S.L. Luxury Car Center - Jaguar Main Page
    - (95) Jaguar Underground

- Entertainment & Media
  - SubCategory
  - More (14)
    - (93) Tom's Collection of Jaguar Mark II Photos
    - (93) Macworld Jaguar Page
    - (90) The Jaguar Photo Gallery

- Travel & Vacations
  - SubCategory
  - More (5)
    - (92) Classic Car Source -- Welsh Jaguar Classic Car Museum

- Business & Finance
  - SubCategory
  - More (2)
    - (56) Jaguar Consulting, Inc.

- Shopping & Services
  - SubCategory
  - More (2)

**List Interface**

- Search Results
  - More (80)
    - (99) H.D. Rogers & Sons Auto Parts Jaguar MG Triumph Renault Peugeot
      - Ferrari Fiat B
    - (90) Jaguar Club of North America
    - (97) Welcome to Jaguar
    - (96) Terry's Jaguar Parts
    - (95) Clan Jaguar Quake & Quake 2 Clan
    - (94) Jaguar Club of Florida
    - (93) Jaguar Daimler Heritage Trust
    - (92) Classic Car Source -- Welsh Jaguar Classic Car Museum
    - (91) A.S.L. Luxury Car Center - Jaguar Main Page
    - (90) Jaguar Underground
    - (88) Jaguar Owners Web Ring
    - (87) Jaguar, new cars, used cars, and accessories
    - (96) Jaguar Sovereign
    - (92) Eager Jaguar, your specialist in luxury foreign sports cars and Jaguar automobi
    - (94) A.S.L. Luxury Car Center - Jaguar Main Page
    - (93) Tom's Collection of Jaguar Mark II Photos
    - (90) Macworld Jaguar Page
    - (97) Welcome to Jaguar House1
    - (90) The Jaguar Photo Gallery
User Study

Dumais, S., E. Cutrell, and H. Chen. Optimizing search by showing results in context, CHI 2001
**Subjective Results**

**Dumais, S., E. Cutrell, and H. Chen. Optimizing search by showing results in context, CHI 2001**

7-point rating scale (1=disagree; 7=agree)

<table>
<thead>
<tr>
<th>Question</th>
<th>Category</th>
<th>List</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>It was easy to use this software.</td>
<td>6.4</td>
<td>3.9</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>I liked using this software</td>
<td>6.7</td>
<td>4.3</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>I prefer this to my usual Web Search engine</td>
<td>6.4</td>
<td>4.3</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>It was easy to get a good sense of the range of alternatives</td>
<td>6.4</td>
<td>4.2</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>I was confident that I could find information if it was there.</td>
<td>6.3</td>
<td>4.4</td>
<td>p&lt;.001</td>
</tr>
</tbody>
</table>

The "More" button was useful                                     | 6.5      | 6.1  | n.s.         |
The display of summaries was useful                              | 6.5      | 6.4  | n.s.         |

**Average Number of Uses of Feature per Task**

<table>
<thead>
<tr>
<th>Interface Features</th>
<th>Category</th>
<th>List</th>
<th>significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expanding / Collapsing Structure</td>
<td>0.78</td>
<td>0.48</td>
<td>p&lt;.003</td>
</tr>
<tr>
<td>Viewing Summaries in Tooltips</td>
<td>2.99</td>
<td>4.60</td>
<td>p&lt;.001</td>
</tr>
<tr>
<td>Viewing Web Pages</td>
<td>1.23</td>
<td>1.41</td>
<td>p&lt;.053</td>
</tr>
</tbody>
</table>
Results: Search Time

Category: 56 secs
List: 85 secs  $p < .002$

50% faster with Category interface

Top20: 57 secs
NotTop20: 98 secs

No reliable interaction between query difficulty and interface condition

Category interface is helpful for both easy and difficult queries
Faceted Navigation (Flamenco)

Marti Hearst, SUI 2009

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Clustering Search Results

Marti Hearst, SUI 2009

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Lecture 5 Plan

- Generating result summaries (abstracts)
  - Beyond result list

- Spelling correction and query suggestion

- New directions in search user interfaces
  - Collaborative Search
  - Collaborative Question Answering

- PhD studies in the U.S.
Query Spelling Correction

We have included grand copthorne hotel results - Show only grand copthorn hotel

Grand Cophorne Hotel Singapore
Five-star hotel in the city with guaranteed great rates for online...
www.asiatravel.com

Luxury Hotel Singapore | Official Site Grand Cophorne Waterfront Hotel ...
Luxury hotel Singapore, experience the great comfort and the relaxing environment that matches your life style at Grand Cophorne Waterfront Hotel nestled next to ...

Millennium & Copthorne Hotels
... Official Site of Millennium and Copthorne Hotels offering a magnitude of exceptional ...
Copthorne Hotels plc (LSE: ML.C) is a dynamic global hotel company...
Reformulations from Bad to Good Spellings

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-rewrite</td>
<td>mic amps  -&gt; create taxi</td>
<td>53.2%</td>
</tr>
<tr>
<td>insertions</td>
<td>game codes  -&gt; video game codes</td>
<td>9.1%</td>
</tr>
<tr>
<td>substitutions</td>
<td>john wayne bust  -&gt; john wayne statue</td>
<td>8.7%</td>
</tr>
<tr>
<td>deletions</td>
<td>skateboarding pics  -&gt; skateboarding</td>
<td>5.0%</td>
</tr>
<tr>
<td>spell correction</td>
<td>real eastate  -&gt; real estate</td>
<td>7.0%</td>
</tr>
<tr>
<td>mixture</td>
<td>huston's restaurant  -&gt; houston's</td>
<td>6.2%</td>
</tr>
<tr>
<td>specialization</td>
<td>jobs  -&gt; marine employment</td>
<td>4.6%</td>
</tr>
<tr>
<td>generalization</td>
<td>gm reabtes  -&gt; show me all the current auto rebates</td>
<td>3.2%</td>
</tr>
<tr>
<td>other</td>
<td>thansgiving  -&gt; dia de acconde gracias</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

[Jones & Fain, 2003]
Spelling Correction: Noisy Channel Model

Platonic concept of query

Correct Spelling

Typing quickly
Distracted
Forgot how to spell

Typos/spelling errors

Reconstruct original query by “reversing this process”
Modeling Errors

\[ P(q_{\text{correct}} \mid q_{\text{error}}) = p(q_{\text{error}} \mid q_{\text{correct}}) p(q_{\text{correct}}) \]

Error model
Character level: \( p(m\mid n) \) \( p(s\mid z) \) etc

Language Model
Query level: \( p(\text{“sigir 2008”}) \), \( p(\text{“sigir iraq”}) \)...

Mine web data sources for these probabilities

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Learning Spell Checker from Query Logs

[Cucerzan and Brill, 2004]
Spelling Correction: Iterative Approach

[Cucerzan and Brill, 2004]

• Main idea:
  – Iteratively transform the query into other strings that correspond to more likely queries.
  – Use statistics from query logs to determine likelihood.
    • Despite the fact that many of these are misspelled
    • Assume that the less wrong a misspelling is, the more frequent it is, and correct > incorrect

• Example:
  – ditroitigers ->
    • detroitigers ->
      – detroit tigers

<table>
<thead>
<tr>
<th>Query</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>albert einstein</td>
<td>4834</td>
</tr>
<tr>
<td>albert einstien</td>
<td>525</td>
</tr>
<tr>
<td>albert einstine</td>
<td>149</td>
</tr>
<tr>
<td>albert einsten</td>
<td>27</td>
</tr>
<tr>
<td>albert einsteins</td>
<td>25</td>
</tr>
<tr>
<td>albert einstain</td>
<td>11</td>
</tr>
<tr>
<td>albert einthin</td>
<td>10</td>
</tr>
<tr>
<td>albert eintin</td>
<td>9</td>
</tr>
<tr>
<td>albeart einstein</td>
<td>6</td>
</tr>
<tr>
<td>aolbert einstein</td>
<td>6</td>
</tr>
<tr>
<td>alber einstein</td>
<td>4</td>
</tr>
<tr>
<td>albert einseint</td>
<td>3</td>
</tr>
<tr>
<td>albert einsteirn</td>
<td>3</td>
</tr>
<tr>
<td>albert einterin</td>
<td>3</td>
</tr>
<tr>
<td>albert einint</td>
<td>3</td>
</tr>
<tr>
<td>alberto einstein</td>
<td>3</td>
</tr>
<tr>
<td>albrecht einstein</td>
<td>3</td>
</tr>
<tr>
<td>alvert einstein</td>
<td>3</td>
</tr>
</tbody>
</table>
Spelling Correction Algorithm

[Cucerzan and Brill, 2004]

• Compute the set of all possible alternatives for each word in the query
  – Stats on word unigrams, bigrams from logs
  – Handles word concatenation and splitting
• Find the best possible alternative string to the input
  – Use modified Viterbi algorithm
• Constraints:
  – No 2 adjacent in-vocabulary words can change simultaneously
  – Short queries have further (unstated) restrictions
  – In-vocabulary words can’t be changed in the first round of iteration

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Spelling Correction Algorithm (cont’d)

[Cucerzan and Brill, 2004]

• Comparing string similarity
  – Damerau-Levenshtein edit distance:
    • The minimum number of point changes required to transform a string into another

• Trading off distance function leniency:
  – A rule that allows only one letter change can’t fix:
    • dondal duck -> donald duck
  – A too permissive rule makes too many errors:
    • log wood -> dog food

• Actual measure:
  – “a modified context-dependent weighted Damerau-Levenshtein edit function”
    • Point changes: insertion, deletion, substitution, immediate transpositions, long-distance movement of letters
    • “Weights interactively refined using statistics from query logs”
Spelling Correction Evaluation
[Cucerzan and Brill, 2004]

• Emphasizing recall
• First evaluation:
  – 1044 randomly chosen queries
  – Annotated by two people (91.3% agreement)
  – 180 misspelled; annotators provided corrections
  – 81.1% system agreement with annotators
    • 131 false positives
      – 2002 kawasaki ninja zx6e \(\rightarrow\) 2002 kawasaki ninja zx6r
    • 156 suggestions for the misspelled queries
  – 2 iterations were sufficient for most corrections
  – **Problem: annotators were guessing user intent**
Spelling Correction Evaluation
[Cucerzan and Brill, 2004]

• Second evaluation:
  – Try to find a misspelling followed by its correction
    • Sample successive pairs of queries from the log
      – Must be sent by same user
      – Differ from one another by a small edit distance
    • Present the pair to human annotators for verification and placement into the gold standard
      – Paper doesn’t say how many total
Spelling Correction Results

[Cucerzan and Brill, 2004]

• Results on 2nd evaluation:
  – 73.1% accuracy
  – Disagreed with gold standard 99 times; 80 suggestions
    • 40 of these were bad
    • 15 functionally equivalent (audio file vs. audio files)
    • 17 different valid suggestions (phone listings vs. telephone listings)
    • 8 found errors in the gold standard (brandy sniffers)
  – 85.5% correct: speller correct or reasonable
  – Sent an unspecified subset of the errors to Google’s spellchecker
    • Its agreement with the gold standard was slightly lower
General Query Suggestion
[Slides adapted from Jones et al., 2006]
Query Substitutions
[Slides adapted from Jones et al., 2006]

1. **Yahoo! Answers - Sleeping with a dead cat??**
   8 answers - One of my friends father had a cat that had cancer and was slowly dying. They went out to dinner and returned home and the cat was dead. The dad was really close with the cat and had been hand feeding...
   answers.yahoo.com/question/?qid=1006021205038 - 39k - Cached - More from this site - Save

2. **I Love Cats: Faces of the Cat Photo Gallery by Don Northup at pbase.com**
   I Love Cats: Faces of the Cat. Lindsay Landis. 15-Apr-2006 20:06. I love CATS there so cute! I love Your pictures and your websitel Its very cool! manxy 26 - 14-Apr-2006 16:13. I hv the cat pics!!! they are so cute!
Query Substitutions

[Slides adapted from Jones et al., 2006]
Functions of Rewriting

[Slides adapted from Jones et al., 2006]

- Enhance meaning
  - Spell correction
  - Corpus-appropriate terminology
    - Cat cancer → feline cancer
- Change meaning
  - Narrow
    - [lexical entailment: fruit → apple]
  - Broaden
    - [alternatives, common interests]
    - Conference proceedings → textbooks
Example: Trying to Find Nathan Welsh, who lives and works in Edinburgh

[Slides adapted from Jones et al., 2006]

- nathan welsh edinburg scotland
- nathan welsh edinburgh scotland
- financial consultants edinburgh scotland
- financial consultants edinburgh scotland
- financial consultants
- nathan welsh 16-18 pennwell place edinburgh
- nathan welsh 16-18 pennywell place edinburgh
- international phone directory
- white pages
- edinburgh scotland phone directory
- edinburgh scotland uk
- nathan welsh investment consultant edinburgh
- nathan welsh investment consultant edinburgh
- investment consultants edinburgh scotland
- nathan welsh
- kansas virginia
- herndon virginia

Switch to new topic
## Half of Query Pairs are Related

[Slides adapted from Jones et al., 2006]

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-rewrite</td>
<td>mic amps → create taxi</td>
<td>53.2%</td>
</tr>
<tr>
<td>insertions</td>
<td>game codes → video game codes</td>
<td>9.1%</td>
</tr>
<tr>
<td>substitutions</td>
<td>john wayne bust → john wayne statue</td>
<td>8.7%</td>
</tr>
<tr>
<td>deletions</td>
<td>skateboarding pics → skateboarding</td>
<td>5.0%</td>
</tr>
<tr>
<td>spell correction</td>
<td>real eastate → real estate</td>
<td>7.0%</td>
</tr>
<tr>
<td>mixture</td>
<td>huston's restaurant → houston's</td>
<td>6.2%</td>
</tr>
<tr>
<td>specialization</td>
<td>jobs → marine employment</td>
<td>4.6%</td>
</tr>
<tr>
<td>generalization</td>
<td>gm reabtes → show me all the current auto rebates</td>
<td>3.2%</td>
</tr>
<tr>
<td>other</td>
<td>thanksgiving → dia de acconde gracias</td>
<td>2.4%</td>
</tr>
</tbody>
</table>

[Jones & Fain SIGIR 2003]
Substitutions are repeated
[Slides adapted from Jones et al., 2006]

• car insurance → auto insurance
  – 5086 times in a sample
• car insurance → car insurance quotes
  – 4826 times
• car insurance → geico [ brand of car insurance ]
  – 2613 times
• car insurance → progressive auto insurance
  – 1677 times
• car insurance → carinsurance
  – 428 times

Different Users, Different Days
Statistical Test to Find Significant Rewrites
[Slides adapted from Jones et al., 2006]

Test whether

\[ p(q_2 \mid q_1) \gg p(q_2) \]

\[
\Pr(\text{britney spears} \mid \text{britney spears}) \gg \Pr(\text{britney spears})
\]

8% \gg 0.01%

Log likelihood ratio test (GLRT) gives \( \chi^2 \) distributed score

About 90% of query pairs are related after filtering with LLR > 100
## Many Types of Substitutable Rewrites

[Slides adapted from Jones et al., 2006]

<table>
<thead>
<tr>
<th>Rewrite</th>
<th>Frequency</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog -&gt; dogs</td>
<td>9185</td>
<td>pluralization</td>
</tr>
<tr>
<td>dog -&gt; cat</td>
<td>5942</td>
<td>both instances of 'pet'</td>
</tr>
<tr>
<td>dog -&gt; dog breeds</td>
<td>5567</td>
<td>generalization</td>
</tr>
<tr>
<td>dog -&gt; dog pictures</td>
<td>5292</td>
<td>more specific</td>
</tr>
<tr>
<td>dog -&gt; 80</td>
<td>2420</td>
<td>random junk in query processing</td>
</tr>
<tr>
<td>dog -&gt; pets</td>
<td>1719</td>
<td>generalization -- hypernym</td>
</tr>
<tr>
<td>dog -&gt; puppy</td>
<td>1553</td>
<td>specification -- hyponym</td>
</tr>
<tr>
<td>dog -&gt; dog picture</td>
<td>1416</td>
<td>more specific</td>
</tr>
<tr>
<td>dog -&gt; animals</td>
<td>1363</td>
<td>generalization -- hypernym</td>
</tr>
<tr>
<td>dog -&gt; pet</td>
<td>920</td>
<td>generalization -- hypernym</td>
</tr>
</tbody>
</table>
Increase Tail Coverage with Query Segmentation

[Slides adapted from Jones et al., 2006]

- Query segmented using high mutual information terms
- Most frequent queries: replace whole query
- Infrequent queries: replace constituent phrases

```
<table>
<thead>
<tr>
<th>castles</th>
<th>in</th>
<th>Edinburgh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>medieval castles</td>
<td>near</td>
<td>Glasgow</td>
</tr>
</tbody>
</table>
```

Represents initial query
Represent rewrite query
## Defining Query Relatedness for Sponsored Search

[Slides adapted from Jones et al., 2006]

<table>
<thead>
<tr>
<th>1- Precise Match</th>
<th>A near-certain match. <em>E.g.: automotive insurance - automobile insurance</em>;</th>
</tr>
</thead>
<tbody>
<tr>
<td>2- Approximate Match</td>
<td>A probable, but inexact match with user intent. <em>E.g.: apple music player - ipod shuffle</em></td>
</tr>
<tr>
<td>3- Marginal Match</td>
<td>A distant, but plausible match to a related topic. <em>E.g.: glasses - contact lenses</em></td>
</tr>
<tr>
<td>4- Mismatch</td>
<td>A clear mismatch.</td>
</tr>
</tbody>
</table>

Call \{1,2\} Precise and \{1,2,3\} Broad
Generating Query Substitutions
[Slides adapted from Jones et al., 2006]

- Q1 $\rightarrow$ \{q2,q3,q4,q5,q6\}
- “catholic baby names” $\rightarrow$
  \{christian baby names, christian baby boy names, catholic names, ...\}
- Learn model to rank and score
Increase Tail Coverage with Query Segmentation

[Slides adapted from Jones et al., 2006]

- Query segmented using high mutual information terms
- Most frequent queries: replace whole query
- Infrequent queries: replace constituent phrases
Generating Query Substitutions

[Slides adapted from Jones et al., 2006]

- Q1 -> \{q2,q3,q4,q5,q6\}
- "catholic baby names" -> \{christian baby names, christian baby boy names, catholic names, ...\}
- All are statistically relevant (log likelihood ratio on successive queries)

Find a model to

- rank substitutions, to be able to pick the best ones
  \[ \text{score}(Q \rightarrow u_1 u_2) < \text{score}(Q \rightarrow Q') < ... \]
- associate a probability of correctness

\[ P(Q \rightarrow Q' \text{ is correct} \mid \text{score}(Q \rightarrow Q')) \]
Train/Test Data
[Slides adapted from Jones et al., 2006]

• Sample 1000 queries (q1)
• Select a single substitution for each (q2)
• Manually label the <q1,q2> pairs
• Learn to score <q1,q2> pairs
• Order by score
• Assess Precision/Recall
  – Precise task \{1,2\} vs \{3,4\}
  – Broad task \{1,2,3\} vs \{4\}
Predicting High Quality Query Suggestions
[Slides adapted from Jones et al., 2006]

- Used labels to fit model
- Tried 37 features for model:
  - Lexical features including
    - Levenshtein character edit distance
    - Prefix overlap
    - Porter-stem
    - Jaccard score on words
  - Statistical features including
    - Probability of rewrite
    - Frequency of rewrite
  - Other
    - Number of substitutions (numSubst)
      - Whole query = 0
      - Replace one phrase = 1
      - Replace two phrases = 2
    - Query length, existence of sponsored results...
Simple Decision Tree

[Slides adapted from Jones et al., 2006]

wordsInCommon > 0

Yes

No

Class={1,2}

prefixOverlap>0

Yes

No

Class={1,2}

Class={3,4}

Interpretation of the decision tree:
• substitution must have at least 1 word in common with initial query
• the beginning of the query should stay unchanged
Linear Regression Model
[Slides adapted from Jones et al., 2006]

**Regression:** continuous output in [1,4]

\[ LMScore = \text{intercept} + \sum_{f=\text{features}} w_f \cdot f \]

**Classification:**

If \((LMScore < T)\) then *Good*, else *Bad*

For each \(T\), we have a precision and a recall

**Evaluation:**

*Average precision / recall on 100 times 10-fold cross validation*
Learned Function
[Slides adapted from Jones et al., 2006]

\[ f(q_1, q_2) = 0.74 + 1.88 \times \text{editDist}(q_1, q_2) \]
\[ + 0.71 \times \text{wordDist}(q_1, q_2) \]
\[ + 0.36 \times \text{numSubst}(q_1, q_2) \]

• Outputs continuous score [1..4]
• Like decision tree
  – Prefer few edits
  – Prefer few word changes
  – Prefer whole-query or few phrase changes
• Normalize output to a probability of correctness using sigmoid fit
SVM, Bags of Trees, Linear Model Trade-offs

[Slides adapted from Jones et al., 2006]
Example Query Substitutions

[Slides adapted from Jones et al., 2006]

<table>
<thead>
<tr>
<th>Initial Query</th>
<th>Substitution</th>
<th>Hand-label</th>
<th>Alg. Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>anne klien watches</td>
<td>anne klein watches</td>
<td>1</td>
<td>92%</td>
</tr>
<tr>
<td>sea world san diego</td>
<td>sea world san diego tickets</td>
<td>2</td>
<td>90%</td>
</tr>
<tr>
<td>restaurants in washington dc</td>
<td>restaurants in washington</td>
<td>2</td>
<td>89%</td>
</tr>
<tr>
<td>nash county</td>
<td>wilson county</td>
<td>3</td>
<td>66%</td>
</tr>
<tr>
<td>frank sinatra birth certificate</td>
<td>elvis presley birth</td>
<td>4</td>
<td>17%</td>
</tr>
</tbody>
</table>
Lecture 5 Plan

- Generating result summaries (abstracts)
  - Beyond result list

- Spelling correction and query suggestion

- New directions in search user interfaces
  - Collaborative Search
  - Collaborative Question Answering

- PhD studies in the U.S.
Collaborative Web Search

- People collaborate during Web search (Morris, 2008)

- Tools have been developed to support collaborative Web search (Morris, 2007; Pickens et al., 2008)

- Information seeking can be more effective as a collaboration than as a solitary activity.
  - Different perspectives, experiences, expertise, and vocabulary to the search process.
Algorithmically Mediated Social Search

• Previous approaches (above): merge searching results from different individuals or let multiple people share a single user interface and cooperatively formulate queries

• **Pickens et al.**: algorithmically-mediated retrieval in search engine level to focus and enhance the team’s search and communication activities

Algorithmically Mediated Social Search II

J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back. 
Algorithmic mediation for collaborative exploratory search, SIGIR 2008

• Two search roles:
  – Prospector: opens new fields for exploration into a data collection.
  – Miner: view and assess the documents returned by Prospector.

• System architecture
  – User Interface Layer
    • A query interface for Prospector to issue queries.
    • A visualization result browsing interface for Miner to assess relevance.
  – Regulator Layer
    • Input regulator is responsible for capturing and storing searcher’s searching results.
    • Output regulator accepts information from the algorithmic layer and routes it to appropriate roles.
• Algorithmic Layer
  – Weight Definition
    • $L_k$: a ranked list of documents retrieved by query $k$.
    • Relevance: $w_r(L_k) = |\text{rel} \in L_k| / |\text{nonrel} \in L_k|$
    • Freshness: $w_f(L_k) = |\text{unseen} \in L_k| / |\text{seen} \in L_k|$
  – Miner Algorithm
    • As Prospector generates new search results, new list ($L_k$) is added to the whole results collection ($L$).
    • The documents retrieved by Prospector will be queued for Miner to assess their relevance. The queue is ordered by the following formula in which $\text{borda}(d, L_k)$ is a function to measure the importance of document $d$ in $L_k$:
      \[
      \text{score}(d) = \sum_{L_k \in \{L\}} w_r(L_k)w_f(L_k)\text{borda}(d, L_k)
      \]
    • Both Prospector and Miner will view and judge documents, so the weights ($w_f$ and $w_r$) will change over time.
    • As a result, the documents with higher scores will have more chances to be evaluated by the Miner.
• **Prospector Algorithm**
  
  — Prospector focuses on coming up with new avenues for exploration into the collection. This is accomplished by real-time query term suggestion.
  
  — Each term in the whole document corpus has a score which is defined by the following formula. $rlf()$ function means the number of documents in $L_k$ in which term $t$ is found.

\[
\text{score}(t) = \sum_{L_k \in \{L\}} w_r(L_k)w_f(L_k)rlf(t, L_k)
\]

  — As Miner’s algorithm affect $w_f$ and $w_r$, the system will reorder term suggestions.
    
    • The more the Miner digs into fresher and more relevant documents, the more terms associated with those documents will appear in term suggestion.
    • Once one document proves to be not fresh and relevant, the associated terms will be gradually replaced by others.

• **Collaboration is accomplished by the dynamic change of freshness value and relevance value.**
Experimental Setup

- Goal: test the hypothesis that mediated collaboration search offers more effective searching capability than simple merging of independently produced results.

- 4 teams, each team has 2 persons. Every time, one team searches for one topic in two ways:
  - simple merging and mediated collaboration search. Each experiment lasts 15 min.

- 24 topics from TREC collection into two groups based on the total number of relevant documents available for that topic.
  - Topics that fell below the median (130) were deemed “sparse” (average of 60 relevant documents per topic).
  - Topics above the median were “plentiful” (average of 332 relevant documents per topic).
  - Searching “sparse” topics is an exploratory search process, more difficult.
## Results

J. Pickens, G. Golovchinsky, C. Shah, P. Qvarfordt, and M. Back.  
*Algorithmic mediation for collaborative exploratory search*, SIGIR 2008

<table>
<thead>
<tr>
<th></th>
<th>3.75 min</th>
<th>7.5 min</th>
<th>11.25 min</th>
<th>15 min</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Avg%Chg</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_g )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>+9.8</td>
<td>+21.5</td>
<td>+22.4</td>
<td>+30.2</td>
</tr>
<tr>
<td>Plentiful</td>
<td>-2.6</td>
<td>+6.1</td>
<td>+4.2</td>
<td>+0.4</td>
</tr>
<tr>
<td>Sparse</td>
<td>+22.4</td>
<td>+36.8</td>
<td>+40.7</td>
<td>+60.1</td>
</tr>
<tr>
<td>( R_g )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>+15.2</td>
<td>+35.7</td>
<td>+19.2</td>
<td>+29.7</td>
</tr>
<tr>
<td>Plentiful</td>
<td>+13.9</td>
<td>+13.5</td>
<td>+3.8</td>
<td>-4.4</td>
</tr>
<tr>
<td>Sparse</td>
<td>+16.4</td>
<td>+57.9</td>
<td>+34.7</td>
<td>+63.8</td>
</tr>
<tr>
<td>( P_v )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>+13.6</td>
<td>+65.4</td>
<td>+41.1</td>
<td>+51.1</td>
</tr>
<tr>
<td>Plentiful</td>
<td>+16.6</td>
<td>+9.1</td>
<td>+2.3</td>
<td>-9.7</td>
</tr>
<tr>
<td>Sparse</td>
<td>+10.6</td>
<td>+121.6</td>
<td>+79.9</td>
<td>+111.9</td>
</tr>
</tbody>
</table>
Lecture 5 Plan

✓ Generating result summaries (abstracts)
  ✓ Beyond result list

✓ Spelling correction and query suggestion

➢ New directions in search user interfaces
  – Collaborative Search
  ➢ Collaborative Question Answering

• PhD studies in the U.S.
Anthony

I have an unweighted 3.73 GPA on a 4.0 scale, and a weighted 3.82. I've only taken a couple honors classes throughout high school (Chemistry and Math 9) and no APs, but I'm taking two APs this year (senior year) (Economics and Psychology). I've taken the ACTs twice and scored a 29 Composite with a 9 out of 12 on the writing my first time, and a 30 Composite with a 9 on the essay on my second time. I'm a pretty well-rounded student as I have been on missions trips to 3

Ranto

Best Answer - Chosen by Voters

Your GPA is average for Emory. However, the average Emory student has more AP classes than you do. You are on the right track taking more -- but you aren't there yet.

Your ACT score corresponds to an SAT score of about 1920-1980. Over 75% of those who are accepted at Emory have higher SAT scores.

Bottom line -- you are close to where you should be and have a shot at Emory -- but I would put your odds at less than 50%. While I think you have a decent shot at getting into Emory, I think it is pretty unlikely that you will get in Early Decision when your stats are below the average for students who are admitted.

If you take the SATs and score above 2100, then you have a better chance.

You will also need a killer admissions essay.

1 year ago

Source(s):
College Professor
Finding Information Online (Revisited)

Next generation of search:
Algorithmically-mediated information exchange

CQA (collaborative question answering):
• Realistic information exchange
• Searching archives
• Train NLP, IR, QA systems
• Study of social behavior, norms

Content quality, asker satisfaction

Current and future work
Finding High Quality Content in SM

E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, *Finding High Quality Content in Social Media, in WSDM 2008*

- Well-written
- Interesting
- Relevant (answer)
- Factually correct
- Popular?
- Provocative?
- Useful?

As judged by professional editors
Do I have a shot at Emory University?

I have an unweighted 3.73 GPA on a 4.0 scale, and a weighted 3.82. I've only taken a couple honors classes throughout high school (Chemistry and Math 9) and no APs, but I'm taking two APs this year (senior year) (Economics and Psychology). I've taken the ACTs twice and scored a 29 Composite with a 9 out of 12 on the

---

Best Answer - Chosen by Voters

Your GPA is average for Emory. However, the average Emory student has more AP classes than you do. You are on the right track taking more -- but you aren't there yet.

Your ACT score corresponds to an SAT score of about 1920-1980. Over 75% of those who are accepted at Emory have higher SAT scores.
Text analysis

Readability statistics

Language modeling

Punctuation density

Help! math! histogram! asap?
🌟 In Mathematics - Asked by Malikme123 - 0 answers - 3 minutes ago

Capitalization errors

WHAT is heidi montag thinking WITH THIS MUSIC VIDEO?
🌟 In Celebrities - Asked by chris_bann88 - 0 answers - 3 minutes ago

Number of words

Help!!!!!!!!!!!!!!!
🌟 In General - Asked by 'So Confused' - 1 answer - 6 minutes ago

+ spacing density, syllables per word,...
Language model disagreement

Distributions of word n-grams and part-of-speech sequences

when|how|why -- “to” -- verb
“how to identify …”
when|how|why -- verb -- verb -- pronoun -- verb
“how do I remove …”
Link Analysis for Authority Estimation

\[ A(j) = \sum_{i=0..M} H(i) \]
\[ H(i) = \sum_{j=0..K} A(j) \]
Do I have a shot at Emory University?

I have an unweighted 3.73 GPA on a 4.0 scale, and a weighted 3.82. I've only taken a couple honors classes throughout high school (Chemistry and Math 9) and no APs, but I'm taking two APs this year (senior year) (Economics and Psychology). I've taken the ACTs twice and scored a 29 Composite with a 9 out of 12 on the

Best Answer - Chosen by Voters
Your GPA is average for Emory. However, the average Emory student has more AP classes than you do. You are on the right track taking more -- but you aren't there yet.

Your ACT score corresponds to an SAT score of about 1920-1980. Over 75% of those who are accepted at Emory have higher SAT scores.
Yahoo! Answers: The Good News

- Active community of millions of users in many countries and languages

- Effective for **subjective** information needs
  - Great forum for socialization/chat

- Can be invaluable for hard-to-find information not available on the web
Yahoo! Answers: The Bad News

May have to wait a long time to get a satisfactory answer

1. FIFA World Cup
2. Optical
3. Poetry
4. Football (American)
5. Soccer
6. Medicine
7. Winter Sports
8. Special Education
9. General Health Care
10. Outdoor Recreation

May never obtain a satisfying answer
Predicting Asker Satisfaction
Y. Liu, J. Bian, and E. Agichtein, in SIGIR 2008

Given a question submitted by an asker in CQA, predict whether the user will be satisfied with the answers contributed by the community.

— “Satisfied”:

• The asker has closed the question AND
• Selected the best answer AND
• Rated best answer >= 3 “stars” (# not important)

— Else, “Unsatisfied”
# Satisfaction by Topic

<table>
<thead>
<tr>
<th>Topic</th>
<th>Questions</th>
<th>Answers</th>
<th>A per Q</th>
<th>Satisfied</th>
<th>Asker rating</th>
<th>Time to close by asker</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 FIFA World Cup</td>
<td>1194</td>
<td>35,659</td>
<td>329.86</td>
<td>55.4%</td>
<td>2.63</td>
<td>47 minutes</td>
</tr>
<tr>
<td>Mental Health</td>
<td>151</td>
<td>1159</td>
<td>7.68</td>
<td>70.9%</td>
<td>4.30</td>
<td>1.5 days</td>
</tr>
<tr>
<td>Mathematics</td>
<td>651</td>
<td>2329</td>
<td>3.58</td>
<td>44.5%</td>
<td>4.48</td>
<td>33 minutes</td>
</tr>
<tr>
<td>Diet &amp; Fitness</td>
<td>450</td>
<td>2436</td>
<td>5.41</td>
<td>68.4%</td>
<td>4.30</td>
<td>1.5 days</td>
</tr>
</tbody>
</table>
Satisfaction Prediction: Human Judges

• Truth: asker’s rating
• A random sample of 130 questions
• Researchers
  – Agreement: 0.82  F1: 0.45 \(2P\times R/(P+R)\)

• Amazon Mechanical Turk
  – Five workers per question.
  – Agreement: 0.9  F1: 0.61
  – Best when at least 4 out of 5 raters agree
## Performance: ASP vs. Humans (F1, Satisfied)

<table>
<thead>
<tr>
<th>Classifier</th>
<th>F1</th>
<th>Satisfied</th>
<th>ASP NB</th>
<th>0.61</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP SVM</td>
<td>0.69</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASP C4.5</td>
<td>0.75</td>
<td>0.76</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>ASP RandomForest</td>
<td>0.70</td>
<td>0.74</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>ASP Boosting</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td></td>
</tr>
</tbody>
</table>

**Best Human Perf**: 0.61

Human F1 is lower than the random baseline!

ASP is significantly more effective than humans.
Top Features by Information Gain

- 0.14  Q: Askers’ previous rating
- 0.14  Q: Average past rating by asker
- 0.10  UH: Member since (interval)
- 0.05  UH: Average # answers for by past Q
- 0.05  UH: Previous Q resolved for the asker
- 0.04  CA: Average asker rating for category
- 0.04  UH: Total number of answers received

...
Current Work (in Progress)

- Partially supervised reinforcement models of expertise (Bian et al., WWW 2009)
- Real-time CQA
- Sentiment, temporal sensitivity analysis
- Mining forum post for health informatics (disease co-morbidity, drug side-effects, ...)

Lecture 5 Plan

✓ Generating result summaries (abstracts)
  ✓ Beyond result list

✓ Spelling correction and query suggestion

✓ New directions in search user interfaces
  ✓ Collaborative Search
  ✓ Collaborative Question Answering

➢ PhD studies in the U.S.
PhD Studies in the U.S.

• Variants:
  – BS/BA (4-years) → MS (2 years) → PhD (4-6 years, 5 year MLE)
  – BS/BA (4-years) → MS + PhD (4-7 years, 5 year MLE)

• Application process
  – Deadline: Late Dec → Mid January
  – Standard Exam Scores:
    • GRE general
    • TOEFL
  – Application:
    • Personal statement/research interests
    • Reference letters
    • Transcript (grades).

• Other resources:
  – Pavel Dmitriev page:
    http://www.pavel-dmitriev.org/faq/question001_ru.xml
Emory Intelligent Information Access Lab (IRLab) (we are hiring…)

- Text and data mining
- Modeling information seeking behavior
- Web search and social media search
- Tools for medical informatics and public health

In collaboration with:
- Beth Buffalo (Neurology)
- Charlie Clarke (Waterloo)
- Ernie Garcia (Radiology)
- Phil Wolff (Psychology)
- Hongyuan Zha (GaTech)

Ablimit Aji (2nd year PhD)  
Qi Guo (3rd year PhD)

1st year graduate students: Julia Kiseleva, Dmitry Lagun, Qiaoling Liu, Wang Yu
Online Behavior and Interactions

Information sharing: blogs, forums, discussions

Search logs: queries, clicks

Client-side behavior: Gaze tracking, mouse movement, scrolling
Research Overview

Discover Models of Behavior
(machine learning/data mining)

Intelligent search
Information sharing
Health Informatics
Cognitive Diagnostics

Eugene Agichtein, RuSSIR 2009, September 11-15, Petrozavodsk, Russia
Main Application Areas

- **Search**: ranking, evaluation, advertising, search interfaces, medical search (clinicians, patients)

- **Collaborative information sharing**: searcher intent, success, expertise, content quality

- **Health informatics**: self reporting of drug side effects, co-morbidity, outreach/education

- **Automatic cognitive diagnostics**: stress, frustration, other impairments ...
References and Further Reading


**Dumais**, S, E. Cutrell, and H. Chen. *Optimizing search by showing results in context*, CHI 2001

**Cucerzan**, S and Brill, E, *Spelling Correction as an Iterative Process that Exploits the Collective Knowledge of Web Users*, EMNLP 2004

**Jones**, R., Rey, B., Madani, O., and Greiner, W. *Generating query substitutions*, WWW 2006
