CS 572: Information Retrieval

Implicit Searcher Feedback
Information Retrieval Process Overview

1. Query Formulation
   - Query: car safety ratings
   - Resource selection
   - Source reselection
   - Query reformulation, vocabulary learning, relevance feedback

2. Search
   - Search Engine Result Page (SERP)
   - Documents
   - Query reformulation, vocabulary learning, relevance feedback

3. Selection
   - Ranked List
   - Examination
   - Delivery

Credit: Jimmy Lin, Doug Oard, ...
Implicit Feedback

• Users often reluctant to provide relevance judgments
  – Some searches are precision-oriented (no “more like this”)
  – They’re busy or annoyed:
  – “Was this document helpful?”

• Can we gather relevance feedback without requiring the user to do anything?

• Goal: estimate relevance from behavior

• Problem: need search intent to interpret behavior
• **Wandering:** the user does not have an information seeking-goal in mind. May have a meta-goal (e.g. “find a topic for my final paper.”)

• **Exploring:** the user has a general goal (e.g. “learn about the history of communication technology”) but not a plan for how to achieve it.

• **Seeking:** the user has started to identify information needs that must be satisfied (e.g. “find out about the role of the telegraph in communication.”), but the needs are open-ended.

• **Asking:** the user has a very specific information need that corresponds to a closed-class question (“when was the telegraph invented?”).
User intent taxonomy (Broder 2002)

- **Informational** – want to learn about something (~40% / 65%)
  - History nonya food

- **Navigational** – want to go to that page (~25% / 15%)
  - Singapore Airlines

- **Transactional** – want to do something (web-mediated) (~35% / 20%)
  - Jakarta weather
  - Kalimantan satellite images
  - Nikon Finepix
  - Access a service
  - Downloads
  - Shop

- Gray areas
  - Find a good hub
  - Car rental Kuala Lumpur
  - Exploratory search “see what’s there”
Extended User Goal Taxonomy

Goal category (GHI)

1. navigational
   2.1. informational / directed / closed
   2.1.2. informational / directed / open
   2.2. informational / undirected
   2.3. informational / advice
   2.4. informational / locate
   2.5. informational / list
   3.1. resource / download
   3.2. resource / entertainment
   3.3. resource / interact
   3.4. resource / obtain
   4. other
   5. indeterminate
   6. non-english
   7. non-standard

Resource total

Navigational

Directed

Undirected

Advice

Locate

List

Set 1

Informational total

Download

Entertain

Interact

Obtain

Kelly Blue Book

View search results from AltaVista

View search results from Google

Submit classification

Rose et al., 2004
Goal: maximize rate of information gain.

Patches of information → websites

Basic Problem: should I continue in the current patch or look for another patch?

Expected gain from continuing in current patch, how long to continue searching in that patch

Information Foraging Theory

Pirolli and Card, CHI 1995
Charnov’s Marginal Value Theorem

Diminishing returns: 80% of users scan only first 3 pages of search results.
Goal: Find cheapest 4-star hotel in Paris.

Step 1: pick hotel search site

Step 2: scan list

Step 3: goto 1
Example: Hotel Search (cont’d)
Levels of Understanding Search Behavior

[Daniel M. Russell, 2007]

• Micro (eye tracking): lowest level of detail, milliseconds

• Meso (field studies): mid-level, minutes to days

• Macro (session analysis): millions of observations, days to months
Micro-level: Examining Results

- Users rapidly scan the search result page.
- What they see in lower summaries may influence their judgment of higher results.
- Spend most time scrutinizing top results 1 and 2 – Trust the ranking

[Daniel M. Russell, 2007]
Eye Tracking

- Process of measuring either the point of gaze (where one is looking) or the motion of an eye relative to the head.

- Eye tracker is a device for measuring eye positions and eye movement.

- Used in research on the visual system, in psychology, in cognitive linguistics and in product design.

**Examples of measures:**

- Time to First Fixation
- Fixations Before
- First Fixation Duration
- Fixation Duration
- Total Fixation Duration
- Fixation Count
- Visit Duration
- Visit Count

whole screen or AOI (area of interest)
Macro-Level (Session) Analysis

- Can examine theoretical user models in light of empirical data:
  - Orienteering?
  - Foraging?
  - Multi-tasking?
- Search is often a multi-step process:
  - Find or navigate to a good site ("orienteering")
  - Browse for the answer there: [actor most oscars] vs. [oscars]
- Teleporting
  - “I wouldn’t use Google for this, I would just go to...”
- Triangulation
  - Draw information from multiple sources and interpolate
  - Example: “how long can you last without food?”
Users (sometimes) Multi-task  

[Daniel M. Russell, 2007]  

<table>
<thead>
<tr>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>Google Search [free roulette] (4s) (DUPE) (p=78)</td>
</tr>
<tr>
<td>102</td>
<td>Google Result 7 <a href="http://www.getlyrical.com/general/free_casino_games/free_online_roulette.html">www.getlyrical.com/general/free_casino_games/free_online_roulette.html</a> (3s)</td>
</tr>
<tr>
<td>103</td>
<td>Google Result 7 <a href="http://www.getlyrical.com/general/free_casino_games/free_online_roulette.html">www.getlyrical.com/general/free_casino_games/free_online_roulette.html</a> (19s) (DUPE) (p=100)</td>
</tr>
<tr>
<td>106</td>
<td>Google Result 8 <a href="http://www.saliu.com/Roulette.htm">www.saliu.com/Roulette.htm</a> (56s) (p=100)</td>
</tr>
<tr>
<td>112</td>
<td>Google Search [shockwave] (4s)</td>
</tr>
<tr>
<td>114</td>
<td>Google Result 3 <a href="http://www.shockwave.com/sw/home/">www.shockwave.com/sw/home/</a> (10s)</td>
</tr>
<tr>
<td>117</td>
<td>Google Result 5 sdc.shockwave.com/shockwave/download/download.cgi (16s) (p=112)</td>
</tr>
<tr>
<td>120</td>
<td>Google Search [free roulette] (3s) (DUPE) (p=78)</td>
</tr>
<tr>
<td>122</td>
<td>Google Result 1 <a href="http://www.ildado.com/free_roulette.html">www.ildado.com/free_roulette.html</a> (15s) (DUPE)</td>
</tr>
<tr>
<td>124</td>
<td>Google Search [free professional roulette] (2s)</td>
</tr>
<tr>
<td>126</td>
<td>Google Search (spell correct) [free professional roulette] (10s)</td>
</tr>
<tr>
<td>128</td>
<td>Google Result 3 imagesculptor.com/Roulette/free-roulette-professional-system.php (5s)</td>
</tr>
<tr>
<td>129</td>
<td>Google Result 3 imagesculptor.com/Roulette/free-roulette-professional-system.php (8s) (DUPE) (p=126)</td>
</tr>
<tr>
<td>133</td>
<td>Google Result 7 <a href="http://www.amazon.com/exec/obidos/tg/detail/-/B0007XRSQ4?v=glance">www.amazon.com/exec/obidos/tg/detail/-/B0007XRSQ4?v=glance</a> (2s) (p=126)</td>
</tr>
</tbody>
</table>
Kinds of Search+Browsing Behavior

[Daniel M. Russell, 2007]
Orienteering vs. Teleporting

• **Orienteering:**
  – Searcher issues a quick, imprecise to get to approximately the right information space region
  – Searchers follow known paths that require small steps that move them closer to their goal
  – Easy (does not require to generate a “perfect” query)

• **Teleporting:**
  – Issue (longer) query to jump directly to the target
  – Expert searchers issue longer queries
  – Requires more effort and experience.
  – Until recently, was the dominant IR model

Teevan et al., CHI 2004
Predicting Queries from Browsing Behavior

[Cheng et al., WWW 2010]

- Identify “Search Trigger” browse-search patterns
- Distribution of “Search-Browse” patterns:

<table>
<thead>
<tr>
<th>Category</th>
<th>Number of Queries</th>
<th>Percentage</th>
<th>Average Pattern Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>SearchTrigger</td>
<td>Key phrase of the page</td>
<td>202</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Information in the page but not key phrases</td>
<td></td>
<td>179</td>
</tr>
<tr>
<td>Non-SearchTrigger</td>
<td>Famous site</td>
<td>604</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Unrelated topic</td>
<td></td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>Repeated search</td>
<td></td>
<td>350</td>
</tr>
<tr>
<td></td>
<td>Query refinement</td>
<td></td>
<td>118</td>
</tr>
<tr>
<td>Cannot judge</td>
<td>Cannot judge</td>
<td>43</td>
<td></td>
</tr>
</tbody>
</table>

URLs: movies.about.com/ nationalpriorities.org pds.jpl.nasa.gov/planets

- new movies releases
- recent movies
- dancing movies
- sports movies
- movies of 2006
- cost of iraq war chart
- iraq war
- coalition of the willing members
- war in iraq deaths
- current iraq war debt
- planets
- nasa kids pictures
- saturn photos
- pictures of space shuttle
- ufos
ReFinding Behavior

- 40% of the queries led to a click on a result that the same user had clicked on in a past search session. – Teevan et al., 2007

- What’s the URL for this year’s SIGIR 2010?
  – Does not really matter, it is faster to re-find it
What Is Known About Re-Finding
[From Teevan et al, 2007]

- Re-finding recent topic of interest
- Web re-visitation common [Tauscher & Greenberg]
- People follow known paths for re-finding
  - Search engines likely to be used for re-finding
- Query log analysis of re-finding
  - Query sessions [Jones & Fain]
  - Temporal aspects [Sanderson & Dumais]
# Granularity of Search Behavior

## Minimum Scope

<table>
<thead>
<tr>
<th>Behavior Category</th>
<th>Segment</th>
<th>Object</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examine</td>
<td>View</td>
<td>Select (click)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Listen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retain</td>
<td>Print</td>
<td>Bookmark</td>
<td>Subscribe</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Save</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Purchase</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Delete</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Copy / paste</td>
<td>Forward</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quote</td>
<td>Reply</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Link</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cite</td>
<td></td>
</tr>
<tr>
<td>Annotate</td>
<td>Mark up</td>
<td>Rate</td>
<td>Organize</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publish</td>
<td></td>
</tr>
</tbody>
</table>
People Look at Only a Few Results

“When you perform a search on a search engine and don’t find what you are looking for, at what point do you typically either revise your search, or move on to another search engine? (Select one)”

(Source: iprospect.com WhitePaper_2006_SearchEngineUserBehavior.pdf)
Snippet Views Depend on Rank

Mean: 3.07 Median: 2.00

Total number of abstracts viewed per page

Mean: 3.07 Median: 2.00

Dip after page break
Snippet Views and Clicks Depend on Rank
[from Joachims et al, SIGIR 2005]

Figure 1: Percentage of times an abstract was viewed/clicked depending on the rank of the result.
Clicks as Relevance Feedback

• Limitations:
  – Hard to determine the meaning of a click. If the best result is not displayed, users will click on something
  – Presentation bias
  – Click duration may be misleading
    • People leave machines unattended
    • Opening multiple tabs quickly, then reading them all slowly
    • Multitasking

• Compare above to limitations of explicit feedback:
  – Sparse, inconsistent ratings
“Strawman” Click model: No Bias

[Craswell et al., 2008]

• Naive Baseline

\[ C_{di} = r_d = C_{dj} \]

– cd\(i\) is \(P(\text{Click}=\text{True} \mid \text{Document}=d, \text{Position}=i)\)
– rd is \(P(\text{Click}=\text{True} \mid \text{Document}=d)\)

• Why this baseline?

– We know that rd is part of the explanation
– Perhaps, for ranks 9 vs 10, it’s the main explanation
– It is a bad explanation at rank 1 e.g. Eye tracking

Attractiveness of summary \(\sim=\) Relevance of result
Realistic Click models

• Clickthrough and subsequent browsing behavior of individual users influenced by many factors
  – Relevance of a result to a query
  – Visual appearance and layout
  – Result presentation order
  – Context, history, etc.
Simple Model: Deviation from Expected

- Relevance component: deviation from “expected”:
  \[
  \text{Relevance}(q, d) = \text{observed} - \text{expected} (p)
  \]

[Agichtein et al., 2006]
Simple Model: Example

- CD: distributional model, extends SA+N
  - Clickthrough considered iff frequency $> \varepsilon$ than expected

- Click on result 2 likely “by chance”
- $4>(1,2,3,5), \text{ but not } 2>(1,3)$
Another Formulation

[Craswell et al., 2008]

• There are two types of user/interaction
  – Click based on relevance
  – Click based on rank (blindly)

\[ C_{di} = \lambda r_d + (1 - \lambda) b_i \]

• A.k.a. the OR model:
  – Clicks arise from relevance OR position
  – Estimate with logistic regression
Linear Examination Hypothesis

- Users are less likely to look at lower ranks, therefore less likely to click

\[ C_{di} \equiv R_d \times i \]

- This is the AND model
  - Clicks arise from relevance AND examination
  - Probability of examination \( x_i \) does not depend on what else is in the list

[Craswell et al., 2008]
Cascade Model

• Users examine the results in rank order

• At each document $d$
  – Click with probability $r_d$
  – Or continue with probability $(1-r_d)$

\[
C_d[i] = r_d \prod_{j=1}^{i-1} (1 - r_{\text{doc in rank}:j})
\]
Cascade Model (2)

query

URL1

URL2

URL3

URL4

r1 \rightarrow r2 \rightarrow r3 \rightarrow r4

r_d \rightarrow C_1 \rightarrow C_2 \rightarrow C_3 \rightarrow C_4

Relevance

ClickThroughs

(1-r_d) \rightarrow (1-r_d) \rightarrow (1-r_d)
Cascade Model Example

[Craswell et al., 2008]

- 500 users typed a query
- 0 click on result A in rank 1
- 100 click on result B in rank 2
- 100 click on result C in rank 3

- Cascade (with no smoothing) says:
  - 0 of 500 clicked A $\Rightarrow$ $r_A = 0$
  - 100 of 500 clicked B $\Rightarrow$ $r_B = 0.2$
  - 100 of remaining 400 clicked C $\Rightarrow$ $r_C = 0.25$
Cascade Model Seems Closest to Reality  
[Craswell et al., 2008]

<table>
<thead>
<tr>
<th>Model</th>
<th>Cross Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Possible</td>
<td>0.141 ± 0.0055</td>
</tr>
<tr>
<td>Cascade</td>
<td>0.225 ± 0.0052</td>
</tr>
<tr>
<td>Logistic</td>
<td>0.236 ± 0.0063</td>
</tr>
<tr>
<td>Examination</td>
<td>0.247 ± 0.0072</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.250 ± 0.0073</td>
</tr>
</tbody>
</table>

Best possible: Given the true click counts for ordering BA
Dynamic Bayesian Net

O. Chapelle, & Y Zhang, A Dynamic Bayesian Network Click Model for Web Search Ranking, WWW 2009

- did user examine url?
- was user satisfied by landing page?
- user attracted to url?
Dynamic Bayesian Net

O. Chapelle, & Y Zhang, A Dynamic Bayesian Network Click Model for Web Search Ranking, WWW 2009

- Did user examine url?
- Was user satisfied by landing page?
- User attracted to url?

\[ A_i = 1, E_i = 1 \Leftrightarrow C_i = 1 \]
\[ P(A_i = 1) = a_u \]
\[ P(S_i = 1|C_i = 1) = s_u \]
\[ C_i = 0 \Rightarrow S_i = 0 \]
\[ S_i = 1 \Rightarrow E_{i+1} = 0 \]
\[ P(E_{i+1} = 1|E_i = 1, S_i = 0) = \gamma \]
\[ E_i = 0 \Rightarrow E_{i+1} = 0 \]
Dynamic Bayesian Net

O. Chapelle, & Y Zhang, A Dynamic Bayesian Network Click Model for Web Search Ranking, WWW 2009

\[ A_i = 1, E_i = 1 \Leftrightarrow C_i = 1 \]
\[ P(A_i = 1) = a_u \]
\[ P(S_i = 1|C_i = 1) = s_u \]
\[ C_i = 0 \Rightarrow S_i = 0 \]
\[ S_i = 1 \Rightarrow E_{i+1} = 0 \]
\[ P(E_{i+1} = 1|E_i = 1, S_i = 0) = \gamma \]
\[ E_i = 0 \Rightarrow E_{i+1} = 0 \]

\[ r_u := P(S_i = 1|E_i = 1) \]
\[ = P(S_i = 1|C_i = 1)P(C_i = 1|E_i = 1) \]
\[ = a_u s_u \]
Dynamic Bayesian Net (results)

O. Chapelle, & Y Zhang, A Dynamic Bayesian Network Click Model for Web Search Ranking, WWW 2009

Use EM algorithm (similar to forward-backward to learn model parameter $\gamma$) set manually

predicted relevance agrees 80% with human relevance
Assumption: If a user skips a link $a$ and clicks on a link $b$ ranked lower, then the user preference reflects $\text{rank}(b) < \text{rank}(a)$.

Example: $(3 < 2) \text{ and } (7 < 2), (7 < 4), (7 < 5), (7 < 6)$

**Ranking Presented to User:**
1. Kernel Machines  
   [http://svm.first.gmd.de/](http://svm.first.gmd.de/)
2. Support Vector Machine  
3. SVM-Light Support Vector Machine  
4. An Introduction to Support Vector Machines  
5. Support Vector Machine and Kernel ... References  
6. Archives of SUPPORT-VECTOR-MACHINES ...  
   [http://www.jiscmail.ac.uk/lists/SUPPORT...](http://www.jiscmail.ac.uk/lists/SUPPORT...)
7. Lucent Technologies: SVM demo applet  
8. Royal Holloway Support Vector Machine  
   [http://svm.dcs.rhbnc.ac.uk/](http://svm.dcs.rhbnc.ac.uk/)
**Results**

<table>
<thead>
<tr>
<th>weight</th>
<th>feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.60</td>
<td>cosine between query and abstract</td>
</tr>
<tr>
<td>0.48</td>
<td>ranked in top 10 from Google</td>
</tr>
<tr>
<td>0.24</td>
<td>cosine between query and the words in the URL</td>
</tr>
<tr>
<td>0.24</td>
<td>document was ranked at rank 1 by exactly one of the 5 search engines</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0.17</td>
<td>country code of URL is &quot;.de&quot;</td>
</tr>
<tr>
<td>0.16</td>
<td>ranked top 1 by HotBot</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>-0.15</td>
<td>country code of URL is &quot;.fi&quot;</td>
</tr>
<tr>
<td>-0.17</td>
<td>length of URL in characters</td>
</tr>
<tr>
<td>-0.32</td>
<td>not ranked in top 10 by any of the 5 search engines</td>
</tr>
<tr>
<td>-0.38</td>
<td>not ranked top 1 by any of the 5 search engines</td>
</tr>
</tbody>
</table>
Extension: Query Chains

[Radlinski & Joachims, KDD 2005]

There is extra information in query reformulations.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ithaca Tompkins Regional Airport - Home</td>
</tr>
<tr>
<td>2</td>
<td>Ithaca Tompkins Regional Airport - Schedule</td>
</tr>
<tr>
<td>3</td>
<td>Ithaca Airport, Greece: Travel information</td>
</tr>
</tbody>
</table>

“ithaca airport”

<table>
<thead>
<tr>
<th>Rank</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cornell Remote Sensing</td>
</tr>
<tr>
<td>2</td>
<td>Cheap Flights Ithaca - Discount Airfares</td>
</tr>
<tr>
<td>3</td>
<td>Ithaca Tompkins Regional Airport - The ...</td>
</tr>
</tbody>
</table>

“ithaca airline”

<table>
<thead>
<tr>
<th>Rank</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cornell Remote Sensing</td>
</tr>
<tr>
<td>2</td>
<td>14850 Today: News for Ithaca, New York</td>
</tr>
<tr>
<td>3</td>
<td>14850 Today: Independence Air to serve ...</td>
</tr>
</tbody>
</table>

“new ithaca airline”
Query Chains (Cont’d)

[Radlinski & Joachims, KDD 2005]
Query Chains (Results)

[Radlinski & Joachims, KDD 2005]

- Query Chains add slight improvement over clicks

<table>
<thead>
<tr>
<th>Evaluation Mode</th>
<th>Chains</th>
<th>User Prefers</th>
<th>Other</th>
<th>Indifferent</th>
</tr>
</thead>
<tbody>
<tr>
<td>(rel_{QC} vs. )</td>
<td>392 (32%)</td>
<td>239 (20%)</td>
<td>579 (47%)</td>
<td></td>
</tr>
<tr>
<td>(rel_0)</td>
<td>211 (17%)</td>
<td>160 (13%)</td>
<td>855 (70%)</td>
<td></td>
</tr>
<tr>
<td>(rel_{QC} vs. )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(rel_{NC})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Results on Cornell Library search engine. \(rel_0\) is the original retrieval function, \(rel_{QC}\) is that trained using query chains, and \(rel_{NC}\) is that trained without using query chains.
Recent Papers on LTR from Clicks

• Predicting Relevance from Clicks
• Utility of Clicks vs. Explicit judgments
Problem: Users click based on result “Snippets”

[Clarke et al., 2007]
Clickthrough Inversions

[Clarke et al., 2007]
Relevance is Not the Dominant Factor!  

[Clarke et al., 2007]

<table>
<thead>
<tr>
<th>Relationship</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>rel(A) &lt; rel(B)</td>
<td>119</td>
<td>33.5%</td>
</tr>
<tr>
<td>rel(A) = rel(B)</td>
<td>134</td>
<td>37.7%</td>
</tr>
<tr>
<td>rel(A) &gt; rel(B)</td>
<td>102</td>
<td>28.7%</td>
</tr>
</tbody>
</table>

Figure 3: Relevance relationships at clickthrough inversions. Compares relevance between the higher ranking member of a caption pair (rel(A)) to the relevance of the lower ranking member (rel(B)), where caption A received fewer clicks than caption B.
## Feature Importance

[Clarke et al., 2007]

<table>
<thead>
<tr>
<th>Feature Tag</th>
<th>INV+</th>
<th>INV−</th>
<th>%+</th>
<th>CON+</th>
<th>CON−</th>
<th>%+</th>
<th>(\chi^2)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MissingSnippet</td>
<td>185</td>
<td>121</td>
<td>60.4</td>
<td>144</td>
<td>133</td>
<td>51.9</td>
<td>4.2443</td>
<td>0.0393</td>
</tr>
<tr>
<td>SnippetShort</td>
<td>20</td>
<td>6</td>
<td>76.9</td>
<td>12</td>
<td>16</td>
<td>42.8</td>
<td>6.4803</td>
<td>0.0109</td>
</tr>
<tr>
<td>TermMatchTitle</td>
<td>800</td>
<td>559</td>
<td>58.8</td>
<td>660</td>
<td>700</td>
<td>48.5</td>
<td>29.2154</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>TermMatchTS</td>
<td>310</td>
<td>213</td>
<td>59.2</td>
<td>269</td>
<td>216</td>
<td>55.4</td>
<td>1.4938</td>
<td>0.2216</td>
</tr>
<tr>
<td>TermMatchTSU</td>
<td>236</td>
<td>138</td>
<td>63.1</td>
<td>189</td>
<td>149</td>
<td>55.9</td>
<td>3.8088</td>
<td>0.0509</td>
</tr>
<tr>
<td>TitleStartQuery</td>
<td>1058</td>
<td>933</td>
<td>53.1</td>
<td>916</td>
<td>1096</td>
<td>45.5</td>
<td>23.1999</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>QueryPhraseMatch</td>
<td>465</td>
<td>346</td>
<td>57.3</td>
<td>427</td>
<td>422</td>
<td>50.2</td>
<td>8.2741</td>
<td>0.0040</td>
</tr>
<tr>
<td>MatchAll</td>
<td>8</td>
<td>2</td>
<td>80.0</td>
<td>1</td>
<td>4</td>
<td>20.0</td>
<td>0.0470</td>
<td></td>
</tr>
<tr>
<td>URLQuery</td>
<td>277</td>
<td>188</td>
<td>59.5</td>
<td>159</td>
<td>315</td>
<td>33.5</td>
<td>63.9210</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>URLSlashes</td>
<td>1715</td>
<td>1388</td>
<td>55.2</td>
<td>1380</td>
<td>1758</td>
<td>43.9</td>
<td>79.5819</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>URLLenDiff</td>
<td>2288</td>
<td>2233</td>
<td>50.6</td>
<td>2062</td>
<td>2649</td>
<td>43.7</td>
<td>43.2974</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Official</td>
<td>215</td>
<td>142</td>
<td>60.2</td>
<td>133</td>
<td>215</td>
<td>38.2</td>
<td>34.1397</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Home</td>
<td>62</td>
<td>49</td>
<td>55.8</td>
<td>64</td>
<td>82</td>
<td>43.8</td>
<td>3.6458</td>
<td>0.0562</td>
</tr>
<tr>
<td>Image</td>
<td>391</td>
<td>270</td>
<td>59.1</td>
<td>315</td>
<td>335</td>
<td>48.4</td>
<td>15.0735</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Readable</td>
<td>52</td>
<td>43</td>
<td>54.7</td>
<td>31</td>
<td>48</td>
<td>39.2</td>
<td>4.1518</td>
<td>0.0415</td>
</tr>
</tbody>
</table>
### Important Words in Snippet

[Clarke et al., SIGIR 2007]

<table>
<thead>
<tr>
<th>Rank</th>
<th>Term</th>
<th>$\chi^2$</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>encyclopedia</td>
<td>114.6891</td>
<td>↓</td>
</tr>
<tr>
<td>2</td>
<td>wikipedia</td>
<td>94.0033</td>
<td>↓</td>
</tr>
<tr>
<td>3</td>
<td>official</td>
<td>36.5566</td>
<td>↑</td>
</tr>
<tr>
<td>4</td>
<td>and</td>
<td>28.3349</td>
<td>↑</td>
</tr>
<tr>
<td>5</td>
<td>tourism</td>
<td>25.2003</td>
<td>↑</td>
</tr>
<tr>
<td>6</td>
<td>attractions</td>
<td>24.7283</td>
<td>↑</td>
</tr>
<tr>
<td>7</td>
<td>free</td>
<td>23.6529</td>
<td>↓</td>
</tr>
<tr>
<td>8</td>
<td>sexy</td>
<td>21.9773</td>
<td>↑</td>
</tr>
<tr>
<td>9</td>
<td>medlineplus</td>
<td>19.9726</td>
<td>↓</td>
</tr>
<tr>
<td>10</td>
<td>information</td>
<td>19.9115</td>
<td>↑</td>
</tr>
</tbody>
</table>

**Figure 6:** Words exhibiting the greatest positive (↑) and negative (↓) influence on clickthrough patterns.
Richer Behavior Models

- Behavior measures of Interest
  - Browsing, scrolling, dwell time
  - How to estimate relevance?

- Heuristics

- Learning-based
  - General model: Curious Browser [Fox et al., TOIS 2005]
  - Query + Browsing [Agichtein et al., SIGIR 2006]
  - Active Prediction: [Yun et al., WWW 2010]
Curious Browser

Evaluate the result that you just visited:

- I liked it.
- It was interesting, but I need more information.
- I didn't like it.
- I did not get a chance to evaluate it (broken link, foreign language, etc).

Is this search:
- An entirely new search?
- A refinement of your previous search?

Evaluate your previous search for sigir implicit...
- I was satisfied with the search.
- I was partially satisfied with the search.
- I was not satisfied with the search.

[Fox et al., 2003]

Eugene Agichtein
Emory University
Data Analysis

- Bayesian modeling at result and session level
- Trained on 80% and tested on 20%
- Three levels of SAT – VSAT, PSAT & DSAT
- Implicit measures:

<table>
<thead>
<tr>
<th>Result-Level</th>
<th>Session-Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diff Secs, Duration Secs</td>
<td>Averages of result-level measures (Dwell Time and Position)</td>
</tr>
<tr>
<td>Scrolled, ScrollCnt, AvgSecsBetweenScroll,</td>
<td>Query count</td>
</tr>
<tr>
<td>TotalScrollTime, MaxScroll</td>
<td></td>
</tr>
<tr>
<td>TimeToFirstClick, TimeToFirstScroll</td>
<td>Results set count</td>
</tr>
<tr>
<td>Page, Page Position, Absolute Position</td>
<td>Results visited</td>
</tr>
<tr>
<td>Visits</td>
<td>End action</td>
</tr>
<tr>
<td>Exit Type</td>
<td></td>
</tr>
<tr>
<td>ImageCnt, PageSize, ScriptCnt</td>
<td></td>
</tr>
<tr>
<td>Added to Favorites, Printed</td>
<td></td>
</tr>
</tbody>
</table>
Result-Level Findings

1. Dwell time, clickthrough and exit type strongest predictors of SAT
2. Printing and Adding to Favorites highly predictive of SAT when present
3. Combined measures predict SAT better than clickthrough

[Fox et al., 2003]
Learning Result Preferences in Rich User Interaction Space

[Agichtein et al., 2006]

• Observed and Distributional features
  – Observed features: aggregated values over all user interactions for each query and result pair
  – Distributional features: deviations from the “expected” behavior for the query

• Represent user interactions as vectors in “Behavior Space”
  – **Presentation**: what a user sees *before* click
  – **Clickthrough**: frequency and timing of clicks
  – **Browsing**: what users do *after* the click
Features for Behavior Representation

[Agichtein et al., SIGIR2006]

<table>
<thead>
<tr>
<th>Presentation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ResultPosition</strong></td>
<td><strong>Position of the URL in Current ranking</strong></td>
</tr>
<tr>
<td><strong>QueryTitleOverlap</strong></td>
<td><strong>Fraction of query terms in result Title</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clickthrough</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DeliberationTime</strong></td>
<td><strong>Seconds between query and first click</strong></td>
</tr>
<tr>
<td><strong>ClickFrequency</strong></td>
<td><strong>Fraction of all clicks landing on page</strong></td>
</tr>
<tr>
<td><strong>ClickDeviation</strong></td>
<td><strong>Deviation from expected click frequency</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Browsing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DwellTime</strong></td>
<td><strong>Result page dwell time</strong></td>
</tr>
<tr>
<td><strong>DwellTimeDeviation</strong></td>
<td><strong>Deviation from expected dwell time for query</strong></td>
</tr>
</tbody>
</table>

Sample Behavior Features
Predicting Result Preferences

- Task: predict pairwise preferences
  - A judge will prefer Result A > Result B

- Models for preference prediction
  - Current search engine ranking
  - Clickthrough
  - Full user behavior model

Eugene Agichtein Emory University
User Behavior Model

Full set of interaction features
- Presentation, clickthrough, browsing

Train the model with explicit judgments
- Input: behavior feature vectors for each query-page pair in rated results
  - Use RankNet (Burges et al., [ICML 2005]) to discover model weights
  - Output: a neural net that can assign a “relevance” score to a behavior feature vector
Results: Predicting User Preferences

[Agichtein et al., SIGIR2006]

- Baseline < SA+N < CD << UserBehavior
- Rich user behavior features result in dramatic improvement
Feature Merging: Details

[Agichtein et al., SIGIR2006]

Query: SIGIR, fake results w/ fake feature values

<table>
<thead>
<tr>
<th>Result URL</th>
<th>BM25</th>
<th>PageRank</th>
<th>Clicks</th>
<th>DwellTime</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>sigir2007.org</td>
<td>2.4</td>
<td>0.5</td>
<td>?</td>
<td>?</td>
<td>...</td>
</tr>
<tr>
<td>Sigir2006.org</td>
<td>1.4</td>
<td>1.1</td>
<td>150</td>
<td>145.2</td>
<td>...</td>
</tr>
<tr>
<td>acm.org/sigs/sigir/</td>
<td>1.2</td>
<td>2</td>
<td>60</td>
<td>23.5</td>
<td>...</td>
</tr>
</tbody>
</table>

- **Value scaling:**
  - Binning vs. log-linear vs. linear (e.g., $\mu=0$, $\sigma=1$)

- **Missing Values:**
  - 0? (meaning for normalized feature values s.t. $\mu=0$?)

- **“real-time”: significant architecture/system problems**
Review: NDCG

- Normalized Discounted Cumulative Gain
- Multiple Levels of Relevance

DCG:
- Contribution of $i$th rank position:
  \[
  \frac{2^{y_i} - 1}{\log(i + 1)}
  \]
- Ex: \[\text{has DCG score of}\]
  \[
  \frac{1}{\log(2)} + \frac{3}{\log(3)} + \frac{1}{\log(4)} + \frac{0}{\log(5)} + \frac{1}{\log(6)} \approx 5.45
  \]

NDCG is normalized DCG
- Best possible ranking as score NDCG = 1
Results for Incorporating Behavior into Ranking

[Agichtein et al., SIGIR2006]

<table>
<thead>
<tr>
<th></th>
<th>MAP</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>RN</td>
<td>0.270</td>
<td></td>
</tr>
<tr>
<td>RN+ALL</td>
<td>0.321</td>
<td>0.052 (19.13%)</td>
</tr>
<tr>
<td>BM25</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>BM25+ALL</td>
<td>0.292</td>
<td>0.056 (23.71%)</td>
</tr>
</tbody>
</table>
## Observable Behavior

### Minimum Scope

<table>
<thead>
<tr>
<th>Behavior Category</th>
<th>Segment</th>
<th>Object</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retain</td>
<td>Print</td>
<td>Bookmark, Save, Purchase, Delete</td>
<td>Subscribe</td>
</tr>
<tr>
<td>Reference</td>
<td>Copy / paste, Quote</td>
<td>Forward, Reply, Link, Cite</td>
<td></td>
</tr>
<tr>
<td>Annotate</td>
<td>Mark up</td>
<td>Rate, Publish</td>
<td>Organize</td>
</tr>
</tbody>
</table>
Starting point: Noisy gaze data from the eye tracker.

2. Fixation detection and saccade classification

3. Reading (red) and skimming (yellow) detection line by line

See G. Buscher, A. Dengel, L. van Elst: “Eye Movements as Implicit Relevance Feedback”, in CHI ’08
Three Feedback Methods Compared

[Buscher et al., 2008]

Input: viewed documents

Gaze-Filter

Gaze-Length-Filter

Reading Speed

Baseline

TF x IDF based on read or skimmed passages

Interest(t) x TF x IDF based on length of coherently read text

ReadingScore(t) x TF x IDF based on read vs. skimmed passages containing term t

TF x IDF based on opened entire documents
Eye Tracking-based RF Results

[Buscher et al., 2008]
Searchers might use the mouse to focus reading attention, bookmark promising results, or not at all.

Behavior varies with task difficulty and user expertise.

---

**Oscar Trivia**
38 actors have won 52 Academy Awards for "Best Actor". The winner is **Jack Nicholson**. The actor with the most total Oscar nominations is **Jack Nicholson**. [Similar pages](http://www.seeing-stars.com/Awards/OscarTrivia.shtml)

---

Result Ex. (cont): Predicting Eye-Mouse coordination

Guo & Agichtein, CHI 2010

Actual Eye-Mouse Coordination

Predicted

No Coordination (30%)

Bookmarking (25%)

Eye follows mouse (25%)
Two basic patterns: “Reading” and “Scanning”:

- “Reading”: consuming or verifying when (seemingly) relevant information is found

- “Scanning”: not yet found the relevant information, still in the process of visually searching
Common Viewing Combines Multiple Patterns
[Task: “number of dead pixels to replace a Mac”]

Relevant (dwell time: 70s)

Not Relevant (dwell time: 80s)

Actively move the cursor with pauses → “reading” dominant

Keep the cursor still and scroll → “scanning” dominant
# Post-click Behavior (PCB) Features

<table>
<thead>
<tr>
<th>Feature Group (30)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dwell time (1)</strong></td>
<td>The time of the page view in seconds</td>
</tr>
<tr>
<td><strong>Rank (1)</strong></td>
<td>The search result rank of the document or the rank of the origin (i.e., the landing page) of the search trail</td>
</tr>
<tr>
<td><strong>Cursor (14)</strong></td>
<td>Cursor movement count, frequency, travel distance and range in pixels, speed, min/max coordinates (overall, x and y)</td>
</tr>
<tr>
<td><strong>Scroll (5)</strong></td>
<td>Scroll count, frequency, distance, speed, maximum</td>
</tr>
<tr>
<td><strong>AOI (3)</strong></td>
<td>Cursor active time, count, frequency in the predefined area of interest (AOI)</td>
</tr>
<tr>
<td><strong>Task (6)</strong></td>
<td>Average dwell time, query count, SERP count, click count, clickthrough rate (CTR), task time</td>
</tr>
</tbody>
</table>
PCB Improves Re-Ranking of Visited Pages

- **PCB** and **PCB_User** consistently outperform **DTR** (dwell time)
- Predictions directly usable by a search engine for improving search ranking quality (for other users)

Guo & Agichtein, WWW’12
Guo, Lagun & Agichtein, CIKM’12
Identifying Attention on Page Regions

[Ageev et al., SIGIR 2013]

• Which portion of the page was examined?

• Applications:
  QA passage retrieval  [EMNLP 2013]
  Snippet generation  [SIGIR 2013]
Motivation: Gaze Points to a Perfect Answer

- **Information need**
  - How many pixels must be dead on an iPad 3 before Apple will replace it?

- **Query**
  - [how many dead pixels iPad 3 replace]

- **Area of interest**
  - “iPad: 3 Dead Pixel -> Apple will replace a new LCD for you”
EMU.js: Mouse Tracking with Text Positions

• Problem
  – Mouse cursor positions are given as browser window coordinates
  – Text coordinates (layout) depends on screen resolution and browser version

• Our approach
  – Evaluate window coordinates for each word on client side
## Predicting Fragment Interestingness

- Page divided into overlapping text fragments
- 6 behavior features
- Label: probability of fragment interestingness \((BScore)\)
- ML method: GBRT

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MouseOverTime</td>
<td>Time duration when the mouse cursor was over the text passage</td>
</tr>
<tr>
<td>MouseNearTime</td>
<td>Time duration when the mouse cursor was close to the text passage in the window ((x \pm 100px, y \pm 70px))</td>
</tr>
<tr>
<td>MouseOverEvents</td>
<td>The number of mouse events during \textit{MouseOverTime}</td>
</tr>
<tr>
<td>MouseNearEvents</td>
<td>The number of mouse events during \textit{MouseNearTime}</td>
</tr>
<tr>
<td>DispTime</td>
<td>Time duration when the text passage has been visible in the browser window (depends on scrollbar position)</td>
</tr>
<tr>
<td>DispMiddleTime</td>
<td>Time duration when the text passage was visible in the middle part of the browser window</td>
</tr>
</tbody>
</table>
Behavior-Biased Snippet Generation

• Combine the text-based score TextScore(f) for a candidate fragment with the behavior-based interestingness score BScore(f)

\[ FScore(f) = \lambda \cdot BScore(f) + (1 - \lambda) \cdot TextScore(f) \]

• \( \lambda \) affects coverage and snippet quality
  – Low \( \lambda \) = prioritize text features (baseline)
  – High \( \lambda \) = prioritize behavior signal
Example Improved Snippet

Query: 
Query/Intent: What sports did Britain get from Australia?

List of Australian inventions - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/List_of_Australian_inventions
Australian inventions consisting of products and technology invented in Australia from pre-European-settlement in 1788 to the ... is used in sports broadcasts and provides viewers with spectacular views of events such as motor racing, which are impossible

List of Australian inventions - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/List_of_Australian_inventions
Australian inventions consisting of products and technology invented in Australia from pre-European-settlement in 1788 to the ... Vale near London, Mr. and Mrs. Edward Hirst of Sydney invented the combination polo and lacrosse sport which was first played

Content

Only

Content + Behavior
Extension to Unseen Queries/Documents: Search Trails

[Bilenko and White, WWW 2008]

- Trails start with a search engine query
- Continue until a terminating event
  - Another search
  - Visit to an unrelated site (social networks, webmail)
  - Timeout, browser homepage, browser closing
Probabilistic Model

[Zhai-Lafferty, Lavrenko]

IR via language modeling

\[ Rel(d_i, q) = p(d_i | q) = \sum_{t_j \in q} p(t_j | q) p(d_i | t_j) \]

Query-term distribution gives more mass to rare terms:

\[ p(t_j | q) = \frac{\exp(-p(t_j))}{\sum_{t_k \in q} \exp(-p(t_k))} \]

Term-website weights combine dwell time and counts

\[ f(d_i, t_j) = \sum_{q': t_j \in q', q' \rightarrow d_i} \log(time(q', d_i)) \]

\[ p(d_i | t_j) = \frac{f(d_i, t_j)}{\sum_{d_k \in D} f(d_k, t_j)} \]
Add $\text{Rel}(q, d_i)$ as a feature to RankNet

<table>
<thead>
<tr>
<th></th>
<th>NDCG@1</th>
<th>NDCG@3</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.68</td>
<td>0.69</td>
<td>0.70</td>
</tr>
<tr>
<td>Baseline+Heuristic</td>
<td>0.69</td>
<td>0.70</td>
<td>0.71</td>
</tr>
<tr>
<td>Baseline+Probabilistic</td>
<td>0.70</td>
<td>0.72</td>
<td>0.73</td>
</tr>
<tr>
<td>Baseline+Probabilistic+RW</td>
<td>0.71</td>
<td>0.73</td>
<td>0.74</td>
</tr>
</tbody>
</table>

[Bilenko and White, WWW 2008]
Search behavior models for Touch Screens

Guo et al., SIGIR 2013
Touch Events and Gestures Data

• For each touch event, record:
  – **Action Type** (namely, down, move and up)
  – **Number of touch points**
  – **X and Y coordinates**
  – **Touch pressure** and size

• A **gesture** is a sequence of touch events, starting with “down” and ending with “up”:
  – E.g., pinch/zoom, swipe
Task: In what year did the USA experience its worst drought? What was the average precipitation in the country that year?
• Post-click examination is a mixture of “Scanning” and “Reading”:
  – “Scanning”: visual searching not followed by reading → not relevant
  – “Reading”: consuming or verifying → relevant

• “Scanning”:
  – Fast swiping, searching for relevant page region(s)

• “Reading”:
  – Inactivity, slow swiping, and sometimes zooming

Relevant: self-reported judgments ≥ 3 (“good”) at 5-point scale

Guo et al., SIGIR 2013
User attention concentrated near the center of the viewport

Scroll patterns correlate with content relevance, user satisfaction
Implicit Feedback Summary

Models proposed to simulate searcher click process

- Increasingly sophisticated and theories
- Assume searcher is rational and consistent

But, searchers are not rational or careful:

- Attracted/repelled by simple features of summaries

→ Interpreting implicit ratings (clicks) still open problem. Active research area, including @Emory 😊
Resources

• WWW 2010 Tutorial on search intent & behavior: www.mathcs.emory.edu/~eugene/www2010tutorial/

• Measuring user engagement (WWW 2013 tutorial): www2013.org/program/measuring-user-engagement/

• Web search click modeling (WSDM 2016 tutorial): http://clickmodels.weebly.com/