Personalization and Search

Jaime Teevan
Microsoft Research
Information Retrieval

User Context

Task/Use Context

Domain Context

Query Words

Ranked List

Query Words

Ranked List

Query
Personalization and Search

• Measuring the value of personalization
  – Do people’s notions of relevance vary?
• Understanding the individual
  – How can we model a person’s interests?
• Calculating personal relevance
  – How can we use the model to measure relevance?
• Other ways to personalize search
  – What other aspects can we personalize?
Personalization and Search

• Measuring the value of personalization
  – An example
  – Lots of relevant results ranked low
  – Best group ranking v. individual ranking

• Understanding the individual

• Calculating personal relevance

• Other ways to personalize search
UC Berkeley Home Page
University of California, Berkeley home page. Berkeley is the oldest of the UC campuses and serves as the flagship of California's public university system. Rich in traditions ...
www.berkeley.edu · Cached page

UC Berkeley - Applying to Berkeley
Avoid the brat pack: Website makes raising joyous kids more practical
www.berkeley.edu/applying · 11/15/2007 · Cached page

City of Berkeley Home Page
Information on city services and departments as well as calendar and links.
www.ci.berkeley.ca.us · 11/15/2007 · Cached page

Computer Science Division | EECS at UC Berkeley
Provides information for prospective students, a staff directory, research projects, seminars and conferences.
www.cs.berkeley.edu · Cached page

WELCOME To Berkeley on the Prairie Path Berkeley, IL
Official site for the city. Map, upcoming events, local information, and a message from the mayor.
www.berkeley.il.us · Cached page

Restaurants in the Berkeley Area
Reviews of selected restaurants in the Berkeley area.
www-rc.lbl.gov/Restaurants · Cached page
Relevant Content Ranked Low

![Graph showing relevance ranking]

- **Highly Relevant**
- **Relevant**
- **Irrelevant**

The graph illustrates how relevant content is ranked from highly relevant to irrelevant across different ranks.
Potential for Personalization

![Graph showing DCG vs Number of People]

- **DCG**: The graph plots DCG against the number of people.
- **Number of People**: The x-axis represents the number of people, ranging from 1 to 6.
- **DCG**: The y-axis represents DCG, ranging from 0.6 to 1.0.
- **Individual**: The red line indicates the DCG for individual cases.
Potential for Personalization

Potential for personalization
Potential for Personalization

Potential for personalization

DCG

Number of People

Potential for personalization
Overview

- Measuring the value of personalization
- Understanding the individual
  - Gather information beyond the query
  - Explicit v. implicit
    - Client-side v. server-side
- Calculating personal relevance
- Other ways to personalize search
Learning More Explicitly v. Implicitly

• Explicit
  – User shares more about query intent
  – User shares more about interests
  – Hard to express interests explicitly

Query Words
berkeley restaurants

California or Illinois? Fancy or low key?
Learning More Explicitly v. Implicitly

• Explicit
  – User shares more about query intent
  – User shares more about interests
  – Hard to express interests explicitly

<table>
<thead>
<tr>
<th>Arts</th>
<th>Business</th>
<th>Computers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Games</td>
<td>Health</td>
<td>Home</td>
</tr>
<tr>
<td>Kids and Teens</td>
<td>News</td>
<td>Recreation</td>
</tr>
<tr>
<td>Reference</td>
<td>Regional</td>
<td>Science</td>
</tr>
<tr>
<td>Shopping</td>
<td>Society</td>
<td>Sports</td>
</tr>
</tbody>
</table>

Intellectual property? Rock climbing? Tobacco and guns
Learning More Explicitly v. Implicitly

• Explicit
  – User shares more about query intent
  – User shares more about interests
  – Hard to express interests explicitly

• Implicit
  – Query context inferred
  – Profile inferred about the user
  – Less accurate, needs lots of data
Profile Information

• Behavior-based
  – Click-through
  – Personal PageRank

• Content-based
  – Categories
  – Term vector

→ [computers: 2, microsoft: 1, click: 4, what: 3, tablet: 1]
Profile Information

- Behavior-based
  - Click-through
  - Personal PageRank
- Content-based
  - Categories
  - Term vector

Server information

- Web page index
- Link graph
- Group behavior
Server-Side v. Client-Side Profile

• **Server-side**
  – Pros: Access to rich Web/group information
  – Cons: Personal data stored by someone else

• **Client-side**
  – Pros: Privacy
  – Cons: Need to approximate Web statistics

• **Hybrid solutions**
  – Server sends necessary Web statistics
  – Client sends some profile information to server
Match Individual to Group

- Can use groups of people to get more data
- Back off from individual → group → all
- Collaborative filtering
Overview

• Measuring the value of personalization
• Understanding the individual
• Calculating personal relevance
  – Behavior-based example
  – Content-based example
• Other ways to personalize search
Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs
Content-Based Relevance

• Explicit relevance feedback
  – Mark documents relevant
  – Used to re-weight term frequencies
Content-Based Relevance

Score = \( \sum t f_i \times w_i \)

\( w_i = log \frac{(N)}{(n_i)} \)

\( w_i = log \frac{(r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)} \)
Content-Based Relevance

- Explicit relevance feedback
  - Mark documents relevant
  - Used to re-weight term frequencies
- Lots of information about the user
  - Consider read documents relevant
  - Use to re-weight term frequencies
Content-Based Relevance

Score = Σ tfᵢ * wᵢ

wᵢ = log \frac{(N)}{(nᵢ)}

wᵢ = log \frac{(rᵢ+0.5)(N-nᵢ-R+rᵢ+0.5)}{(nᵢ-rᵢ+0.5)(R-rᵢ+0.5)}

wᵢ = log \frac{(rᵢ+0.5)(N’-n’ᵢ-R+rᵢ+0.5)}{(n’ᵢ-rᵢ+0.5)(R-rᵢ+0.5)}

Where: N’ = N+R, n’ᵢ = nᵢ+rᵢ
Personalization Performance

• Personalized search hard to evaluate
• Mostly small improvements despite big gap
• Identify ambiguous queries
  – Personalize: “berkeley”
  – Don’t personalize: “uc berkeley homepage”
• Identify easily personalized queries
  – Re-finding queries
Other Ways to Personalize

• Measuring the value of personalization
• Understanding the individual
• Calculating personal relevance
• Other ways to personalize search
  – Match expectation for re-finding queries
  – Personalized snippets
Ranking Results for Re-Finding
Ranking Results for Re-Finding
People Don’t Notice Change
People Don’t Notice Change
People Don’t Notice Change
A **winery** is a building or property that produces wine, or a business involved in the production of wine, such as a wine company. Some wine companies own many wineries. Besides wine making equipment ...

en.wikipedia.org/wiki/Winery
Snippets to Support Re-Finding

Query: “winery”

**Winery - Wikipedia, the free encyclopedia**

A **winery** is a building or property that produces wine, or a business involved in the production of wine, such as a wine company. Some wine companies own many wineries. Besides wine making equipment...

[en.wikipedia.org/wiki/Winery](en.wikipedia.org/wiki/Winery)

If the person has visited the page before:

**Winery - Wikipedia, the free encyclopedia**  Last visit: November 14, 2007

New content: It has been suggested that Winery wastewater be merged into this article or section.

[en.wikipedia.org/wiki/Winery](en.wikipedia.org/wiki/Winery)
Interest-Based Snippets

Query: “winery”

**Winery - Wikipedia, the free encyclopedia**

A *winery* is a building or property that produces wine, or a business involved in the production of wine, such as a wine company. Some wine companies own many wineries. Besides wine making equipment ...

[en.wikipedia.org/wiki/Winery](http://en.wikipedia.org/wiki/Winery)

If the person is interested in Maui:

**Winery - Wikipedia, the free encyclopedia**

A *winery* is a building or property that produces wine, or a business involved in the production of wine, such as a wine company... For example, in Maui there is a pineapple *winery*. ...

[en.wikipedia.org/wiki/Winery](http://en.wikipedia.org/wiki/Winery)
Interest-Based Snippets

Query: “winery”

Winery - Wikipedia, the free encyclopedia

A winery is a building or property that produces wine, or a business involved in the production of wine, such as a wine company. Some wine companies own many wineries. Besides wine making equipment...

en.wikipedia.org/wiki/Winery

If the person is interested in Maui:

Winery - Wikipedia, the free encyclopedia

A winery is a building or property that produces wine, or a business involved in the production of wine, such as a wine company... For example, in Maui there is a pineapple winery...

en.wikipedia.org/wiki/Winery
Summary

• Measuring the value of personalization
  – There’s a big gap between group and individual
• Understanding the individual
  – Building a profile, explicit v. implicit
• Calculating personal relevance
  – Relevance feedback, boost click through
• Other ways to personalize search
  – Rank based on expectation, personalized snippets