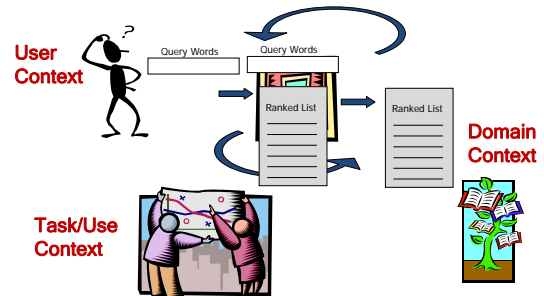


Search Personalization

Jaime Teevan
Microsoft Research

Information Retrieval Query



Personalization and Search

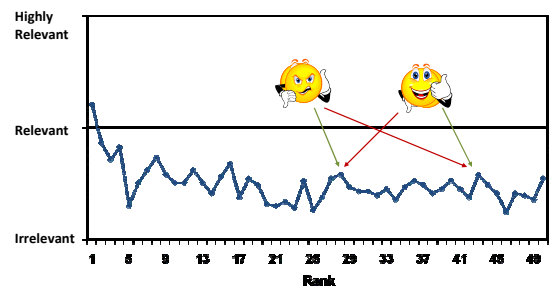
- Measuring the value of personalization
 - Do people’s notions of relevance vary?
- Understanding the individual
 - How can we model a person’s interests?
- Calculating personal relevance
 - How can we use the model to measure relevance?
- Other ways to personalize search
 - What other aspects can we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

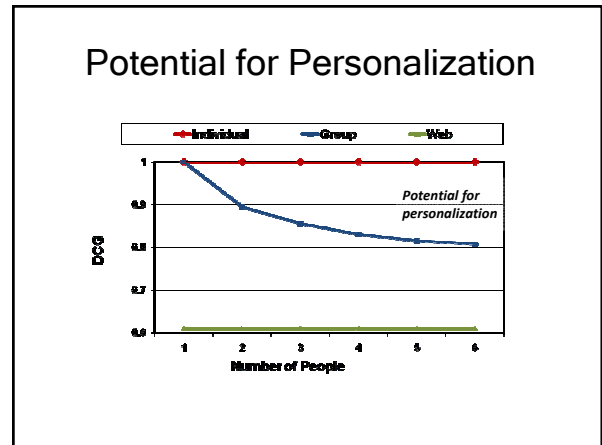
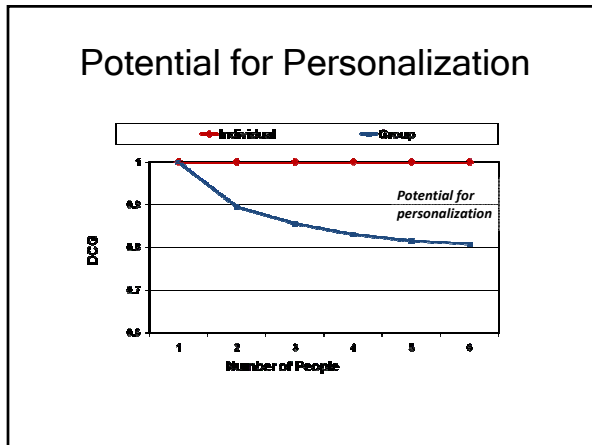
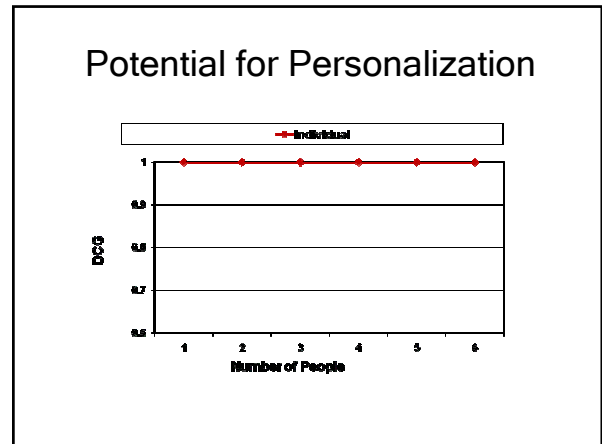
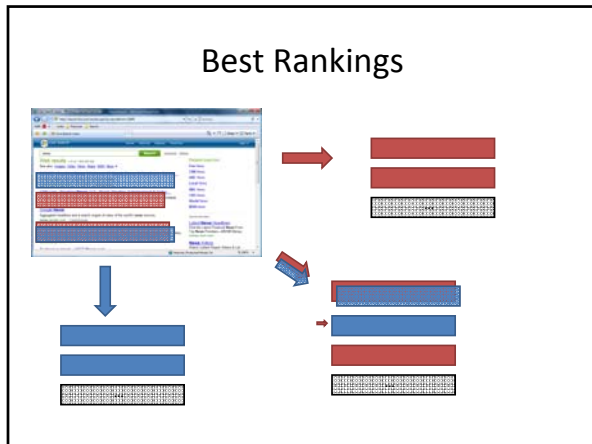
Personalization and Search

- 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



Relevant Content Ranked Low





- ### Overview
- Measuring the value of personalization
 - Understanding the individual
 - Explicit v. implicit
 - Client-side v. server-side
 - Individual v. group
 - 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

uw admissions

Learning More Explicitly v. Implicitly

- Explicit
 - User shares more about query intent
 - User shares more about interests *Intellectual property?*
Rock climbing?
Tobacco and guns
 - Hard to express interests explicitly

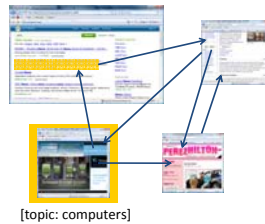
Arts	Business	Computers
Games	Health	Home
Kids and Teens	News	Recreation
Reference	Regional	Science
Shopping	Society	Sports

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



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 or domains

		Repeat Click	New Click	
Repeat Query	33%	29%	4%	43%
New Query	67%	10%	57%	
		39%	61%	

Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs or domains



Behavior-Based Relevance

- People often want to re-find
- People have trusted sites
- Boost previously viewed URLs or domains



Content-Based Relevance

- Explicit relevance feedback
 - Mark documents relevant
 - Used to re-weight term frequencies



Content-Based Relevance

World N

Client n_i

Read R

Intersection r_i

$$\text{Score} = \sum \text{tf}_i * w_i$$

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

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

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

World N

Client n_i

Read R

Intersection r_i

Client n'_i

$$\text{Score} = \sum \text{tf}_i * w_i$$

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

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

$$w_i = \log \frac{(r_i+0.5)(N'-n'_i-R+r_i+0.5)}{(n'_i-r_i+0.5)(R-r_i+0.5)}$$

Where: $N' = N+R$, $n'_i = n_i+r_i$

Personalization Performance

- Personalized search hard to evaluate
- Mostly small improvements despite big gap
- Identify ambiguous queries
 - Personalize: “uw”
 - Don't personalize: “uw seattle library 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



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

If the person has visited the page before:

[Winery - Wikipedia, the free encyclopedia](#) Last visit: November 14, 2007

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

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

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

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