Acknowledgements
Some slides in this lecture adapted from various IR researchers, credits noted on each slide.
Today's Plan

Search Personalization:

- Why?
- Who is a “person”?  
- When to personalize (or not)
- How to personalize?
- How to evaluate?
Personalization is now default


- Bing: http://searchengineland.com/bing-results-get-localized-personalized-64284
personalization research

UMUAI Special Issue: Data Mining for Personalization
Automatic personalization is a central technology used in user adaptive systems to ... upon submission that explains why a briefer discussion of our research ... facweb.cs.depaul.edu/mobasher/umuaidmp

Research Gift Personalization and Engraving from A Gift ... Including unique gift personalization, engraved gifts, gift personalization for her ... department store fare and demand a little more research to find just the right gift. Personalization ... www.giftpersonalization.com/research · Cached page

Which Personalization Tools Work For eCommerce - And ... While Web site personalization has been a hot eCommerce topic for years, managers managers of online businesses know that executing Web site personalization is not easy. In spite of ... www.forrester.com/Research/Document/0,7211,44345,00.html

Personalizing Search via Automated Analysis of ... ... on current Web search. Keywords: Personalized search, personalization, relevance feedback. In: Proceedings of 28th Annual International ACM SIGIR Conference on Research and ... research.microsoft.com/en-us/um/people/horvitz/PS.htm

Personalizing Search
Why Personalize

• How good are search results?
• Do people want the same results for a query?
• How to capture variation in user intent?
  – Explicitly
  – Implicitly
• How can we use what we learn?
Questions

• How good are search results?
• Do people want the same results for a query?
• How to capture variation in user intent?
  – Explicitly
  – Implicitly
• How can we use what we learn?

Teevan et al., TOCHI 2010
How Good Are Search Results?

Teevan et al., TOCHI 2010

Lots of relevant results ranked low

Normalized Gain

Explicit

Rank

0.7

0

1 2 3 4 5 6 7 8 9 10

1 2 3 4 5 6 7 8 9 10

0 1 2 3 4 5 6 7 8 9 10
How Good Are Search Results?

Teevan et al., TOCHI 2010

Behavior data has presentation bias

Lots of relevant results ranked low
How Good Are Search Results?

Behavior data has presentation bias.

Content data also identifies low results.

Lots of relevant results ranked low.

Teevan et al., TOCHI 2010
Do People Want the Same Results?

• What’s best for *personalization research*?
  – For you?
  – For everyone?

• When it’s just you, can rank perfectly

• With many people, ranking must be a compromise
Do People Want the Same Results?

Teevan et al., TOCHI 2010

Potential for Personalization

Normalized DCG

Number of People in Group

Group
Individual

Normalized DCG

0.55
0.7
0.85
1

0.55
0.7
0.85
1

1
2
3
4
5
6

Number of People in Group
How to Capture Variation (1)?

Behavior gap smaller because of presentation bias

Teevan et al., TOCHI 2010
How to Capture Variation (2)?

Teevan et al., TOCHI 2010

- **Normalized DCG**

  - **Explicit**
  - **Behavior**
  - **Content**

**Content data** shows more variation than explicit judgments.

**Behavior gap** smaller because of presentation bias.
What to personalize?

- Identify ambiguous queries
- Solicit more information about need
- Personalize search
  - Using content and behavior-based measures
33% queries repeated
73% of those are navigational

[Teevan et al. SIGIR 2007]
Identifying Personal Navigation

- Repeat queries are often navigational
- The same navigation used over and over again

Was there a unique click on the same result the last 2 times the person issued the query?
Understanding Personal Navigation

- Identified millions of navigation queries
  - Most occur fewer than 25 times in the logs
  - 15% of the query volume

- Queries more ambiguous
  - Rarely contain a URL fragment
  - Click entropy the same as for general Web queries
  - *enquirer* (multiple meanings)
  - *bed bugs* (found navigation)
  - *etsy* (serendipitous encounters)

[Informational]
Etsy.com
Regretsly.com
(parody)
People Express Things Differently

- Differences can be a challenge for Web search
  - *Picture of a man handing over a key.*
  - *Oil painting of the surrender of Breda.*
People Express Things Differently

• Differences can be a challenge for Web search
  – *Picture of a man handing over a key.*
  – *Oil painting of the surrender of Breda.*

• Personalization
  – Closes the gap using more about the person

• *Group*ization
  – Closes the gap using more about the *group*
Person $\sim=$ Group

- Can use groups of people to get more data
- Back off from individual $\rightarrow$ group $\rightarrow$ all
- Collaborative filtering
How to Take Advantage of Groups?

- Who do we share interests with?
- Do we talk about things similarly?
- What algorithms should we use?
Interested in Many Group Types

• Group longevity
  – Task-based
  – Trait-based

• Group identification
  – Explicit
  – Implicit
People Studied

Trait-based dataset
• 110 people
  – Work
  – Interests
  – Demographics
• Microsoft employees

Task-based dataset
• 10 groups x 3 (= 30)
• Know each other
• Have common task
  – “Find economic pros and cons of telecommuting”
  – “Search for information about companies offering learning services to corporate customers”
Queries Studied

Trait-based dataset
• Challenge
  – Overlapping queries
  – Natural motivation
• Queries picked from 12
  – Work
    c# delegates, live meeting
  – Interests
    bread recipes, toilet train dog

Task-based dataset
• Common task
  – Telecommuting v. office
    pros and cons of working in an office
    social comparison
    telecommuting versus office
    telecommuting
    working at home cost benefit
Data Collected

- Queries evaluated
- Explicit relevance judgments
  - 20 - 40 results
  - Personal relevance
    - *Highly relevant*
    - *Relevant*
    - *Not relevant*
- User profile: Desktop index
Answering the Questions

• Who do we share interests with?

• Do we talk about things similarly?

• What algorithms should we use?
Who do we share interests with?

• Variation in query selection
  – Work groups selected similar work queries
  – Social groups selected similar social queries

• Variation in relevance judgments
  – Judgments varied greatly ($\kappa=0.08$)
  – Task-based groups most similar
  – Similar for one query $\neq$ similar for another
Do we talk about things similarly?

- **Group profile similarity**
  - Members more similar to each other than others
  - Most similar for aspects related to the group

<table>
<thead>
<tr>
<th></th>
<th>In task group</th>
<th>Not in group</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>All queries</td>
<td>0.42</td>
<td>0.31</td>
<td>34%</td>
</tr>
<tr>
<td>Group queries</td>
<td>0.77</td>
<td>0.35</td>
<td>120%</td>
</tr>
</tbody>
</table>

- **Clustering profiles recreates groups**
- **Index similarity ≠ judgment similarity**
  - Correlation coefficient of 0.09
What algorithms should we use?

• Calculate personalized score for each member
  – Content: User profile as relevance feedback
    \[
    \sum_{\text{terms}} \frac{tf_i (r_i+0.5)(N-n_i-R+r_i+0.5)}{(n_i-r_i+0.5)(R-r_i+0.5)} \log \text{terms}_i
    \]
  – Behavior: Previously visited URLs and domains
    – [Teevan et al. 2005]
• Sum personalized scores across group
• Produces same ranking for all members
Performance: Task-Based Groups

- Personalization improves on Web
- Groupization gains +5%
Performance: Task-Based Groups

- Personalization improves on Web
- Groupization gains +5%
- Split by query type
  - On-task v. off-task
  - Groupization the same as personalization for off-task queries
  - 11% improvement for on-task queries
Performance: Trait-Based Groups

The graph shows the performance of different groups based on their traits. The Y-axis represents the normalized DCG (Discounted Cumulative Gain), while the X-axis lists various demographic and interest-based groups.

Key observations include:
- **Groupization** and **Personalization** are shown alongside different traits such as gender, age, location, interests, and work-related categories.
- The groups are categorized into interest and work segments.
- The normalized DCG values range from 0.45 to 0.75.
Performance: Trait-Based Groups

Normalized DCG

Interest queries

Work queries

Groupization

Personalization
Performance: Trait-Based Groups

- Trait-Based Groups
- Groupization
- Personalization
- Interest queries
- Work queries
- Interest
- Work

Normalized DCG

- Male
- Female
- Age 20-30
- Age 30-40
- Age 40+
- Seattle
- Seattle suburbs
- Vegetarians
- Photography
- Pets
- Product group I
- Product group II
- Research Group
- Developer
- Program Manager
- Researcher
Local Search (Geographical Personalization)

• Location provides personalization info & context

• *Local search* uses geographic information to modify the ranking of search results
  – location derived from the query text
  – location of the device where the query originated

• e.g.,
  – “underworld 3 cape cod”
  – “underworld 3” from mobile device in Hyannis
Geography and Query Intent

[ Baeza-Yates and Jones] 2008

Location 1: query location
“Pizza Lavista Road, Atlanta, GA”
query1

Distance 1: home-query intent

Location 2: Home address

IP address / profile zip

Location 3: query location

Distance 2: Reformulation distance

“Pizza Emory”

query2
Topic-Distance Profiles

[Baeza-Yates and Jones] 2008

• 20 bins
  – 0 distance
  – Equal fractions of the rest of the data

• Does distribution into distance bins topics vary by topic?

Movie theater  Distant places  Near-by

movie theater  maps  restaurant
How: Approaches

1. Pitkow et al., 2002
2. Qiu et al., 2006
3. Jeh et al., 2003
4. Teevan et al., 2005
5. Das et al., 2007

Figure adapted from: *Personalized search on the world wide web*, by Micarelli, A. and Gasparetti, F. and Sciarrone, F. and Gauch, S., LNCS 2007

Teevan et al., TOCHI 2010
How: System Perspective

1. Pre-retrieval
2. On-retrieval
3. Post-retrieval
Profile Info

- Use of implicit information about the user to create profile and rerank results **locally**.
  - Previously issued queries.
  - Previously visited Web pages.
  - Documents or emails the user has read, created or sent.
What’s in User Profile?

• Behavior-based
  – Click-through
  – Personal PageRank

• Content-based
  – Categories
  – Term vector

→ [computers: 2, microsoft: 1, click: 4, what: 3, tablet: 1]
Profile: Cont’d

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

Server information
- Web page index
- Link graph
- Group behavior
Profile Representation: Details

- For the representation of the user a rich index of personal content was used that captured the user’s interests and computational activities.
- Index included:
  - Web pages.
  - Email messages.
  - Calendar items.
  - Documents stored on the client machine.
- The user representation can be query focused or not.
- Time sensitivity
  - Documents indexed in the last month vs the full index of documents
- User Interests:
  - Query terms issued in the past.
Server-Side vs. 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
Using Profiles: Overview

P. Brusilovsky
Pre-Process: Query Expansion

• User profile is applied to add terms to the query
  – Popular terms could be added to introduce context
  – Similar terms could be added to resolve indexer-user mismatch
  – Related terms could be added to resolve ambiguity
  – Works with any IR model or search engine
Simple Context-based Query Expansion: Chirita et al. 2006

User related documents (desktop documents) containing the query

Score and extract keywords

Top query-dependent, user-biased keywords

Extract query expansion or re-ranking terms

[Chirita, Firan, Nejdl. Summarizing local context to personalize global web search. CIKM 2006]
How: Approaches

1. Pitkow et al., 2002
2. Qiu et al., 2006
3. Jeh et al., 2003
4. Teevan et al., 2005
5. Das et al., 2007

Figure adapted from: Personalized search on the world wide web, by Micarelli, A. and Gasparetti, F. and Sciarrone, F. and Gauch, S., LNCS 2007
PS Search Engine: Teevan et al., SIGIR 2005
PS Search Engine

query

- dog cat
- monkey
- banana
- food
- forest
- hiking
- walking
- baby
- infant
- child
- boy
- girl
- csail mit
- artificial
- research
- robot
- web
- search
- retrieval
- ir
- hunt
PS Search Engine

query

web search retrieval ir hunt

1.3

Search results page
Calculating a Document’s Score

• Based on standard tf.idf

\[
\text{Score} = \sum tf_i \times w_i
\]
Calculating a Document’s Score

• Based on standard tf.idf

\[
\text{Score} = \sum \text{tf}_i \times \text{w}_i
\]
Calculating a Document’s Score

- Based on standard tf.idf

\[
\text{Score} = \sum \text{tf}_i \times \text{w}_i
\]

\[
\text{w}_i = \log \frac{(N)}{(n_i)}
\]
Calculating a Document’s Score

- Based on standard tf.idf

\[ \text{Score} = \sum \text{tf}_i \times w_i \]

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

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

Where: \( N' = N+R, \quad n_i' = n_i+r_i \)

\[ \dagger \text{From Sparck Jones, Walker and Roberson, 1998 [21].} \]
Finding the Parameter Values

• Corpus representation \((N, n_i)\)
  – *How common is the term in general?*
  – Web vs. result set

• User representation \((R, r_i)\)
  – *How well does it represent the user’s interest?*
  – All vs. recent vs. Web vs. queries vs. none

• Document representation
  – *What terms to sum over?*
  – Full document vs. snippet
Building a Test Bed

• 15 evaluators x ~10 queries
  – 131 queries total
• Personally meaningful queries
  – Selected from a list
  – Queries issued earlier (kept diary)
• Evaluate 50 results for each query
  – Highly relevant / relevant / irrelevant
• Index of personal information
Evaluating Personalized Search

• Measure algorithm quality

\[ \text{DCG}(i) = \begin{cases} 
\text{Gain}(i), & \text{if } i = 1 \\
\text{DCG}(i-1) + \frac{\text{Gain}(i)}{\log(i)}, & \text{otherwise}
\end{cases} \]

• Look at one parameter at a time
  
  – 67 different parameter combinations!

  – Hold other parameters constant and vary one

• Look at best parameter combination

  – Compare with various baselines
Analysis of Parameters
Analysis of Parameters

The graph shows the analysis of parameters for different corpus types: Full text, Web, Snippet, None, Query, Web, Recent, All, Snippet, Full text. The y-axis represents the parameter values ranging from 0.27 to 0.35. The x-axis represents the corpus types with categories: Corpus, User, Document.
PS Improves Text Retrieval

- No model
- Relevance Feedback
- Personalized Search
Text Features Not Enough

![Bar chart showing DCG values for different features: No, RF, PS, Web. The DCG values are 0.37, 0.41, 0.46, and 0.56 respectively.]
Take Advantage of Web Ranking

![Bar chart showing DCG values for different combinations: No (0.37), RF (0.41), PS (0.46), Web (0.56), PS+Web (0.58).]
Summary

• Personalization of Web search
  – Result re-ranking
  – User’s documents as relevance feedback

• Rich representations important
  – Rich user profile particularly important
  – Efficiency hacks possible
  – Need to incorporate features beyond text
Much Room for Improvement

• Group ranking
  – Best improves on Web by 23%
  – More people → Less improvement

• Personal ranking
  – Best improves on Web by 38%
  – Remains constant

Potential for Personalization
User Interface Issues

• Make personalization transparent
• Give user control over personalization
  – Slider between Web and personalized results
  – Allows for background computation
• Exacerbates problem with re-finding
  – Results change as user model changes
  – Thesis research – Re:Search Engine
Post-Filter: Annotations

• The result could be relevant to the user in several aspects. Fusing this relevance with query relevance is error prone and leads to a loss of data

• Results are ranked by the query relevance, but annotated with visual cues reflecting other kinds of relevance
  – User interests - Syskill and Webert, group interests - KnowledgeSea
Example: Syskill and Webert

- First example of annotation
- Post-filter to Lycos
- Hot, cold, lukewarm

User Control in Personalization (RF)

TaskSieve: User-Controlled Adaptive Search

1. AFA2000111.1800.0160
   Today, on Saturday, met with about 170 persons were killed in a fire erupted in the electric train elevator at the opening of the season skiing in Austria. The incident was the worst disasters that have occurred in the Alps. It was announced in a statement reported by the governor of Salzburg Franz Austrian news agency that at least 170 persons, mostly young people were killed in a fire erupted on Saturday in an electric train rise() within an expenditure of Austrian Alps. He said the governor of Salzburg to rescue the men were able to lift it came only three bodies on fire on the train as a whole.

2. ZBN2000111.0400.0019
   Salzburg, Austria, Burg introduce the state governor said yesterday that alpine tunnel fire a total of 165 train passengers, of which 12 people escape, and the rest were all killed. Death toll was more than 154 young ski lovers who he said that as a driver were killed in fire, bringing the total number of deaths to 154 people. Tunnel fire extinguished, Austria rescue personnel entered the tunnel began yesterday in the burning of the train to find the bodies of victims.

3. ALH20001112.0700.0037
   Vienna - AFP - The governor of Salzburg Franz announced that at least 172 people, mostly young people, were killed in a fire erupted yesterday in an electric train the rise in one of the Alps Austrian expenditure has not been spared, only six passengers. He added during a press conference at the centre of skiing that eight people only managed to escape from a window, adding that the disaster has led to the killing of. At least 172 people, mostly young people who were skiing with a view to the practice of sport. He said that the lifting of the three men were able to rescue the bodies of the fire came on the train as a whole.

4. AFA20001111.1400.0094
   The governor of Salzburg Franz announced in a statement reported by Austrian news agency that at least 170 persons, mostly young people were killed in a fire erupted on Saturday in an electric train rise() within an Austrian alpine expenditure has not been spared, only eight passengers. He added during a press...
TaskSieve: Adaptive Snippets

• The goal of a usual snippet is to show query relevance. TaskSieve applies adaptive snippets to show profile relevance as well.
• Selects top 3 relevant sentences combining query relevance and task relevance to sentences.
• Applies color coding by query/profile.

Salzburg, Austria, Burg introduce the state governor passengers, of which 12 people escape, and the rest lovers who he said that as a driver were killed in fire, ... Tunnel fire extinguished, Austria rescue personn the train to find the bodies of victims.
Evaluating Personalized Search

- Explicit judgments (offline and in situ)
  - Evaluate components before system
  - NOTE: What’s relevant for you

- Deploy system
  - Verbatim feedback, Questionnaires, etc.
  - Measure behavioral interactions (e.g., click, reformulation, abandonment, etc.)
  - Click biases – order, presentation, etc.
  - Interleaving for unbiased clicks

- Link implicit and explicit (Curious Browser toolbar)
- From single query to search sessions and beyond
Evaluating Personalized Search: Active Area of Research

Evaluating Personal Search Workshop | ECIR2011, 18 April 2011, Dublin, Ireland

Evaluating Personal Search Workshop

[Half day workshop]

Personal Search refers to the process of searching within one's personal space of digital information, e.g., searching one's desktop or mobile phone for required data items or information. This workshop aims to bring together researchers interested in working towards standardized evaluation approaches for the personal search space. Due to the large space that this covers, as a first step towards overall standardized personal search evaluation this workshop will focus on evaluation for the textual elements within personal desktop collections and known item keyword queries for these elements.

News Updates

* Due to numerous requests the paper submission deadline has been extended to February 28, 2011 (midnight Pacific Daylight Time).

Paper Submissions

The workshop is now accepting paper submissions. Short position papers (max. 2 pages) describing creative use of / modifications to provided datasets and approaches or ideas / challenges for the domain are invited. Submissions should be in ACM SIGIR format. LaTeX and Word templates are available at http://www.acm.org/aiga/publications/proceedings-templates (for LaTeX, use the "Option 2" style). Papers should be anonymised and submitted in pdf format through the EasyChair system http://www.easychair.org/conferences/?conf=eps2011 no later than midnight Pacific Daylight Time on February 28, 2011 February 24, 2011. See the workshop call for papers for further details.

Datasets

Jinyoung Kim and Bruce Croft have created simulated desktop collections and user queries. Interested participants can use these resources prior to the workshop in various ways – analyzing the characteristics of the data, evaluating various retrieval methods using the queries and result sets, suggesting a better data set for evaluating personal search, etc. See the datasets page for further information and details on how to obtain the datasets.
Personalization Summary

- Lots of relevant content ranked low
- Potential for personalization high
- Implicit measures capture explicit variation
  - Behavior-based: Highly accurate
  - Content-based: Lots of variation
- Example: Personalized Search
  - Behavior + content work best together
  - Improves search result click through
• Diane Kelly and Jaime Teevan. *Implicit Feedback for Inferring User Preference: A Bibliography*


• Web Spam Taxonomy (paper):