User Information
- We want as much information about the users as possible
  - Demographical
  - Interests and preferences
- But most users are lazy, and are concerned about privacy

Data from User Activities
- Web search queries
- Browsing patterns
  - Links clicked
  - Web pages visited
  - Time spent
- Item purchased
- Posts in bulletin boards, blog
- Ratings, reviews, tags
- ...

Two Topics Related to Web Search
- Understand user goals in web search
- Automatic construction of user profile for personalized search

The Trichotomy of User Search Goals
- Navigational
- Informational
- Transactional

The Search Goal Hierarchy
- Navigational
- Resource
  - Download
  - Entertainment
  - Interact
  - Obtain
- Informational
  - Directed
  - Undirected
  - Advice
  - Locate
  - List

Search Goals Breakdown

A Human Subject Study at UCLA

- 50 most popular Google queries, 28 graduate students
- Query types
  - Clearly navigational, e.g. bestbuy
  - Clearly informational, e.g. hidden markov model
  - Ambiguous, e.g. Alan Kay
- \( i(i) \) – percentage of participants who indicate its goal as informational

Some Results of the Study ...

Some Results of the Study

Automatically Identify Query Goals

- Click distribution
- Anchor-link distribution

Click Distribution Examples
Anchor

Anchor Text

Destination Link

Anchor-link Distribution Examples

Implementation Issues

- Collection of queries and click through data
- Collection of anchor data
- Skewness of the distribution
  - Mean, median, Skewness, Kurtosis
- Combination of features for prediction

Some Thoughts

- Navigational vs. Non-navigational
- Ambiguous and rare queries
- Can we do better??

Personalized Search

- Auto constructed user profile
- Adapting query results

Browsing History

- Days
- Sessions
- Pages

- today
- current session

- ...
Feature Vector of a Page
- Normalized term vector
- A page is ignored if the user didn't spend enough time on it

Session Profile
- Sum of the page feature vectors averaged over number of pages in the session

Ephemeral Profile
- $P^{(br)}$ - average feature vectors of the sessions today before current session
- $P^{(cur)}$ - feature vector of current session
- $P^{\text{today}} = xP^{(br)} + yP^{(cur)}$

Persistent Profile
- $P^{\text{per}}$
- Forgetting factor
- $f = e^{-\log_2(h)(d-d_0)}$
  - $h$: half-life span
  - $d-d_0$: days since the term's last appearance

Profile Based Purely on Browsing History
- $P = aP^{\text{per}} + bP^{\text{today}}$

Profile using Collaborative Filtering
- $P = \begin{array}{cccc}
  \text{user}_1 & \text{user}_2 & \ldots & \text{user}_n \\
  \text{term}_1 \\
  \text{term}_2 \\
  \vdots \\
  \text{term}_m
\end{array}$
Adjust Query Results Based on Profile

\[ \text{sim}(\mathbf{p}, \mathbf{w}) = \frac{\mathbf{p} \cdot \mathbf{w}}{||\mathbf{p}|| \cdot ||\mathbf{w}||} \]

\( \mathbf{p} \): profile
\( \mathbf{w} \): feature vector of a result page

Some Results ...

Summary

◆ Creating user profile without explicit user input
◆ Using user profile to rank items

References

◆ Adaptive Web Search Based on User Profile Constructed without Any Effort from Users, by Sugiyama et. al, 2005.