

CS520 Web Programming

Analyze User Activities

Chengyu Sun
California State University, Los Angeles

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

Andrei Broder, *A taxonomy of Web Search*, 2002.

The Search Goal Hierarchy

- | | |
|-----------------|-----------------|
| ◆ Navigational | ◆ Informational |
| ◆ Resource | ▫ Directed |
| ▫ Download | ▫ Undirected |
| ▫ Entertainment | ▫ Advice |
| ▫ Interact | ▫ Locate |
| ▫ Obtain | ▫ List |

Daniel E. Rose and Danny Levinson,
Understanding User Goals in Web Search, 2002.

Search Goals Breakdown

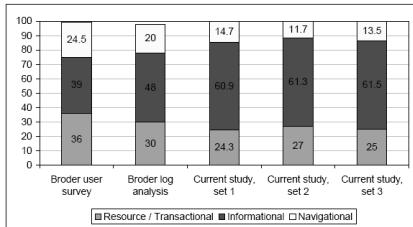


Figure 2: Comparison of Broder's search taxonomy to our top-level goals. Resource and informational results in the first column are Broder's estimates. Results do not total 100% due to rounding error.

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
- ◆ $i(q)$ – percentage of participants who indicate its goal as informational

Some Results of the Study ...

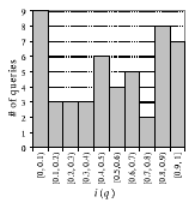


Figure 1: Query distribution along the $i(q)$ axis

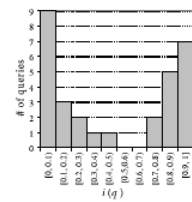


Figure 2: After removing software and person-name queries

... Some Results of the Study

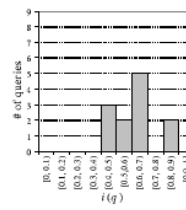


Figure 3: Distribution of the 12 software queries

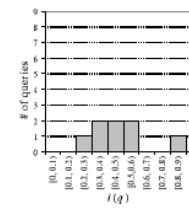


Figure 4: Distribution of the 8 person-name queries

Automatically Identify Query Goals

- ◆ Click distribution
- ◆ Anchor-link distribution

Click Distribution Examples

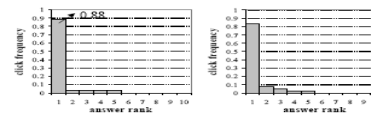


Figure 5: Click distributions for sample navigational queries

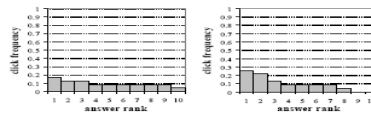
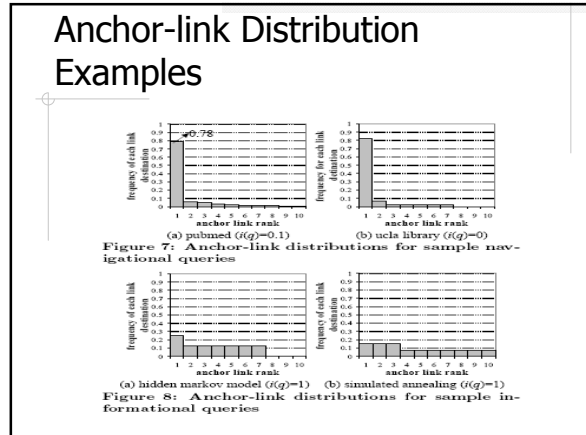
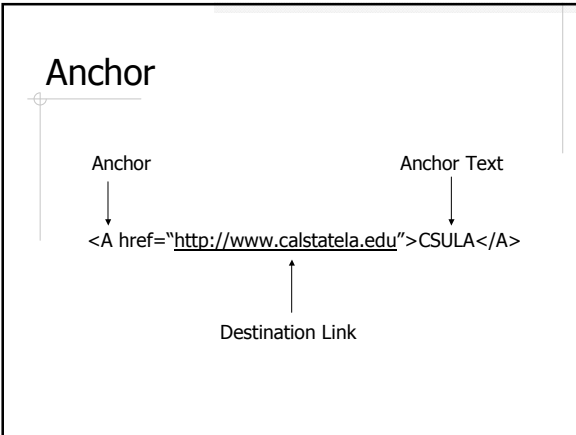


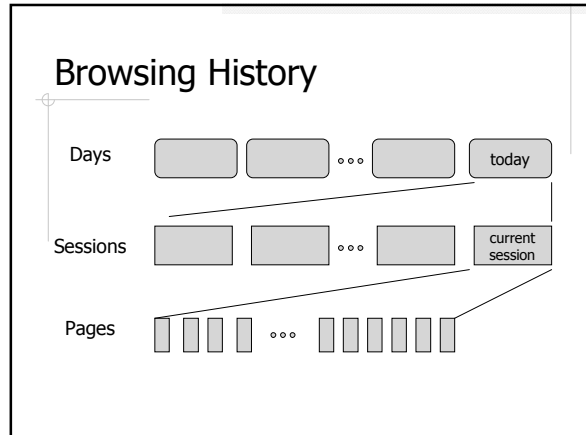
Figure 6: Click distributions for sample informational queries



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

- ## Some Thoughts
- ◆ Navigational vs. Non-navigational
 - ◆ Ambiguous and rare queries
 - ◆ Can we do better??
- n ...
- n ...
- n ...
- n ...

- ## Personalized Search
- ◆ Auto constructed user profile
 - ◆ Adapting query results



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

- ◆ $\mathbf{P}^{(br)}$ - average feature vectors of the sessions today before current session
- ◆ $\mathbf{P}^{(cur)}$ - feature vector of current session
- ◆ $\mathbf{P}^{today} = x\mathbf{P}^{(br)} + y\mathbf{P}^{(cur)}$

Persistent Profile

- ◆ \mathbf{P}^{per}
- ◆ Forgetting factor

$$f = e^{-\frac{\log 2}{h}(d-d_0)}$$

h : half-life span
 $d-d_0$: days since the term's last appearance

Profile Based Purely on Browsing History

$$\mathbf{P} = a\mathbf{P}^{per} + b\mathbf{P}^{today}$$

Profile using Collaborative Filtering

	user ₁	user ₂	...	user _n
term ₁				
term ₂				
.				
.				
.				
term _m				

Adjust Query Results Based on Profile

$$\text{sim}(\mathbf{P}, \mathbf{w}) = \frac{\mathbf{P} \cdot \mathbf{w}}{\|\mathbf{P}\| \|\mathbf{w}\|}$$

\mathbf{P} : profile

\mathbf{w} : feature vector of a result page

Some Results ...

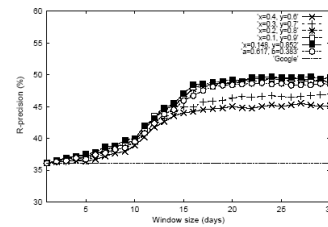


Figure 8: R -precision obtained using pure browsing history-based user profile.

... Some Results ...

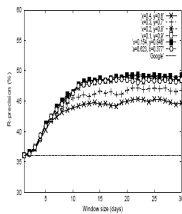


Figure 10: R -precision obtained using modified collaborative filtering-based user profile (static, $n = 10$).

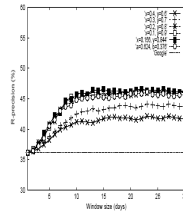


Figure 12: R -precision obtained using modified collaborative filtering-based user profile (static, $n = 20$).

... Some Results

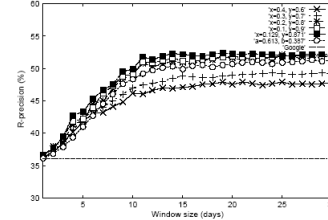


Figure 13: R -precision obtained using modified collaborative filtering-based user profile (dynamic).

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.
- ◆ *Automatic Identification of User Goals in Web Search*, by Lee et. al, 2005.
- ◆ *Understanding User Goals in Web Search*, by Rose et. al, 2004.
- ◆ *A Taxonomy of Web Search* by Broder, 2001.