Understanding and Predicting Personal Navigation

Jaime Teevan, Daniel J. Liebling and Gayathri Ravichandran Geetha

Microsoft Research
33% queries repeated
73% of those are navigational
[Teevan et al. SIGIR 2007]
33% queries repeated
73% of those are navigational
[Teevan et al. SIGIR 2007]
**Road Map of Talk**

- **General Navigation**  
  - Identifying general navigation  
  - Understanding general navigation

- **Personal Navigation**  
  - Identifying personal navigation  
  - Compare with general navigation  
  - Coverage and accuracy of prediction  
  - Consistency of behavior over time

- **Bridging general and personal navigation**

  Bing search logs  
  70 million queries  
  21 million users
Road Map of Talk

- **General Navigation**  *microsoft research*
  - Identifying general navigation
  - Understanding general navigation

- **Personal Navigation**  *wsdm*
  - Identifying personal navigation
  - Compare with general navigation
  - Coverage and accuracy of prediction
  - Consistency of behavior over time

- **Bridging general and personal navigation**

Bing search logs
70 million queries
21 million users
Identifying General Navigation

• Ask people ("Were you looking for this site?")
  – 1 in 4 queries reported to be navigational
• Query string (wsdm.org or microsoft)
  – 10% of queries identified as navigational
• Click behavior
  – Look for low click entropy
  – Need lots of data (query instances, users, clicks)
Understanding General Navigation

• Identified 390 general navigation queries
  – 12% of query volume

• Query strings straightforward
  – facebook, youtube, myspace
  – Short (½ the length of typical Web queries)
  – Contain a URL fragment 20% of the time

• Navigation target usually first result
General Navigation Mistakes

• Click predicted only 72% of the time
  – Double the accuracy for the average query
  – But what’s going on the other 28% of the time?
• Many typical navigation queries not identified
  – craigslist (people visit interior pages)
  – weather.com (people visit related pages)
    3% visit http://geo.craigslist.org/iso/us/ca
    17% visit http://weather.yahoo.com
Road Map of Talk

• General Navigation  *microsoft research*
  – Identify high quality common queries
  – Look navigational ≠ navigational

• Personal Navigation  *wsdm*
  – Identifying personal navigation
  – Compare with general navigation
  – Coverage and accuracy of prediction
  – Consistency of behavior over time

• Bridging general and personal navigation

Bing search logs
70 million queries
21 million users
Road Map of Talk

• General Navigation *microsoft research*
  – Identify high quality common queries
  – Look navigational ≠ navigational

• Personal Navigation *wsdm*
  – Identifying personal navigation
  – Compare with general navigation
  – Coverage and accuracy of prediction
  – Consistency of behavior over time

• Bridging general and personal navigation
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)
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*
    - National Enquirer
    - Cincinnati Enquirer
  - *bed bugs* [Informational]
    - http://www.medicinenet.com/bed_bugs/article.htm
  - *etsy*
    - Etsy.com
    - Regretsy.com (parody)
Personal Navigation Accurate

- Target less likely to be ranked first ..
  - .. than target of general navigation
  - .. than the average Web search click
- Nonetheless, prediction very accurate
  - Correct 95% of the time
Prediction Consistent Over Time

• Looked at different history intervals
  – How much do we need to know about a person?
  – Offline predictions?
• Prediction accuracy consistent over time
• Coverage decreases with stale history
Road Map of Talk

- **General Navigation**  *microsoft research*
  - Identify high quality common queries
  - Look navigational ≠ navigational

- **Personal Navigation**  *wsdm*
  - Re-finding often navigational
  - Identify unusual navigational queries
  - High coverage and accuracy
  - Behavior consistent over time

- **Bridging general and personal navigation**

Bing search logs
70 million queries
21 million users
Road Map of Talk

• General Navigation  *microsoft research*
  – Identify high quality common queries
  – Look navigational ≠ navigational
• Personal Navigation  *wsdm*
  – Re-finding often navigational
  – Identify unusual navigational queries
  – High coverage and accuracy
  – Behavior consistent over time

• Bridging general and personal navigation
Bridging Personal and General

- Some personal navigation queries are general navigation queries

Accuracy of prediction:
- Personal Navigation 95%
- General Navigation 72%

Opportunity to combine aggregate and individual data to increase coverage and drop inaccurate general navigation
Summary of Talk

• General Navigation  *microsoft research*
  – Identify high quality common queries
  – Look navigational ≠ navigational

• Personal Navigation  *wsdm*
  – Re-finding often navigational
  – Identify unusual navigational queries
  – High coverage and accuracy
  – Behavior consistent over time

• General & personal navigation complementary

An opportunity for personalization that works!
Questions?

Jaime Teevan, Daniel J. Liebling and Gayathri Ravichandran Geetha
Microsoft Research