

---

# Query Logs Alone are not Enough

Carrie Grimes, Diane Tang, Dan Russell  
{cgrimes,diane,drussell}@google.com

# Why are query logs so popular?

---

- Lots of data
- Easy to collect: just add code
- Easy to analyze: more code

# BUT...

---

- Many privacy implications
  - Aggregate data can mitigate the implications
  - But everyone wants something different...
- Is the information in query logs worth dealing with the privacy implications? Are there other options?

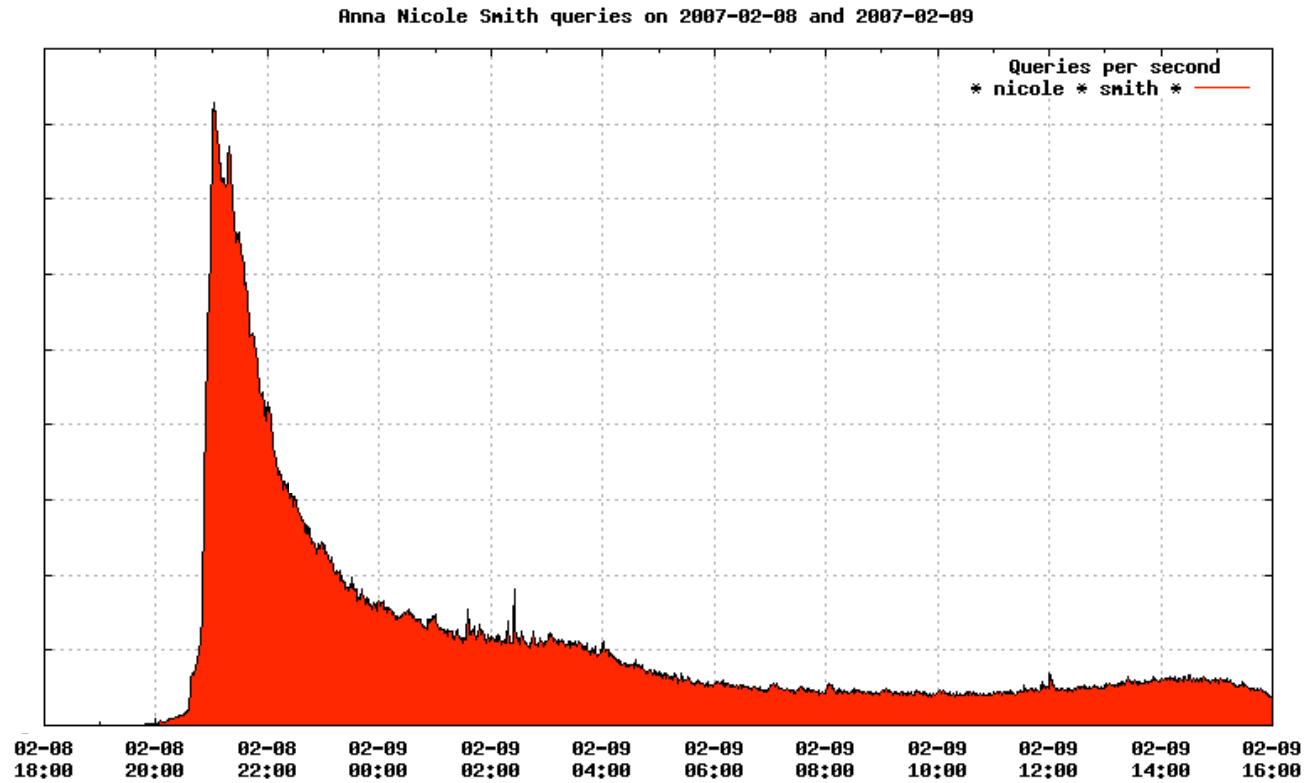
# So what are query logs good for?

---

- Immediacy
- Diversity
- Meaning Disambiguation

# Immediacy / Timeliness

---



Only query logs can give feedback on the order of minutes to detect how interest in a topic shifts or how a query term might change its meaning

# Diversity

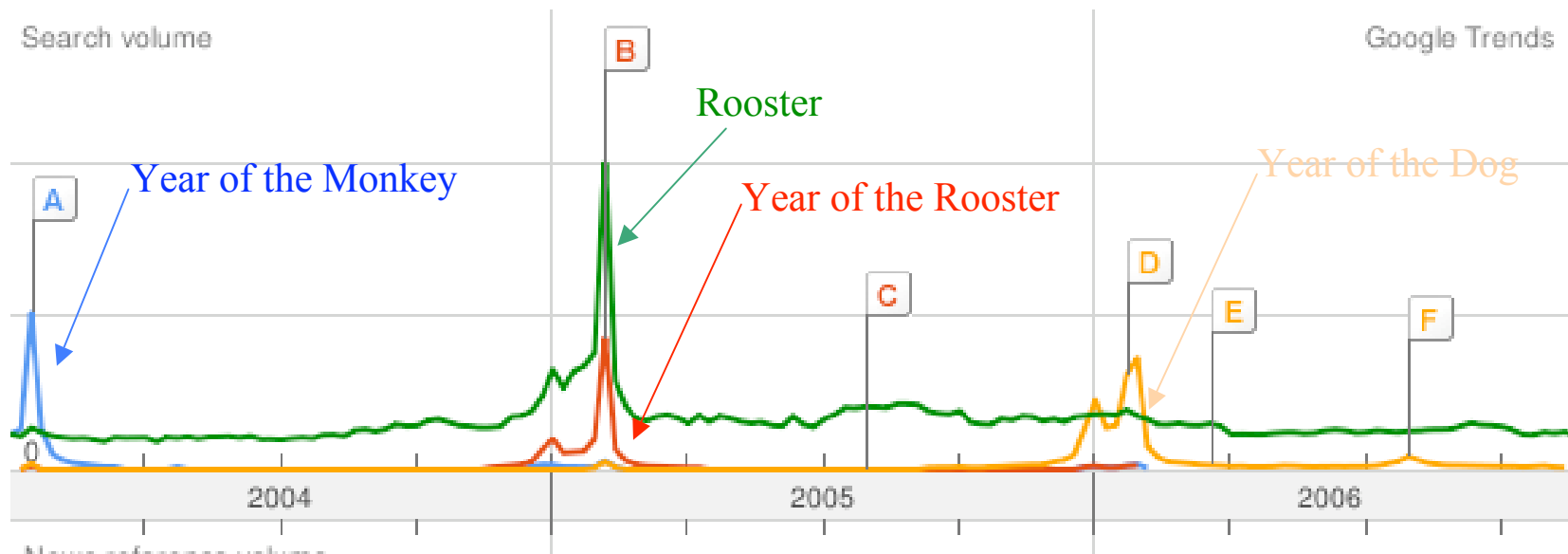
---

- Get diversity from the massive scalability of logs
- Offline, small scale studies have a coverage sample that is too sparse:

# of samples query appeared in out of 25	1 sample	2 samples	3 samples
50 query samples	99.7%	0.2%	0.1%
500 query samples	98.4%	1.1%	0.3%
5000 query samples	96.5%	2.2%	0.6%

- Stratification of the data
  - Effort of covering multiple geographic locales, verticals, task types, etc.
  - A priori vs. post hoc

# Meaning Disambiguation



Aggregate logs: volume of queries can help detect meaning shift for disambiguation, especially in a timely manner.

BUT: intent disambiguation (e.g., shopping vs. research) cannot be gotten from the logs, need alternate data source (e.g., field studies)

# Query logs are actually pretty incomplete!

---

- Completely unlabelled
  - Was a user happy?
  - Where are the task boundaries? Is the user multi-tasking?
  - What was the user actually trying to do?

# What questions are we trying to answer?

---

- What is a search engine trying to do?
  - Do we understand what the user is asking for?
  - Have we presented relevant results for that query?
- Let's focus on the first one...

# Understanding user intent

---

- Task taxonomies
- How users formulate queries
- Types of resources users are looking for

# What was the user looking for?

Event	Load Title
Navigate	Soul Asylum - Google Search
Navigate	SoulAsylum.net ( <a href="http://www.soulasylum.net/">http://www.soulasylum.net/</a> )
Back	Soul Asylum - Google Search
Navigate	Soul Asylum: News ( <a href="http://www.soulasylum.com/">http://www.soulasylum.com/...</a> )
URL Entry	( <a href="http://www.soulasylum.com/">http://www.soulasylum.com</a> )
Navigate	SonyMusicStore: Soul Asylum ( <a href="http://www.sonymusicstore.com/">http://www.sonymusicstore.com/...</a> )
Back	Soul Asylum: News ( <a href="http://www.soulasylum.com/">http://www.soulasylum.com/...</a> )
Back	Soul Asylum - Google Search
Forward	Soul Asylum: News ( <a href="http://www.soulasylum.com/">http://www.soulasylum.com/...</a> )
Back	Soul Asylum - Google Search
Navigate	Soul Asylum - Google Search
Navigate	Soul Asylum's Pirner waiting for word on home - Katrina hits entertainment ( <a href="http://www.MSNBC.com/">http://www.MSNBC.com/...</a> )



\* Data from Google / KeyNote joint study

# What other options are there?

---

- Field / lab studies
- Instrumented panels

# What other options are there?

---

- Field / lab studies
  - 1:1 interaction between researcher and participant
  - Participant performs tasks while researcher watches
  - Observational data (akin to logs) are collected
  - Additional qualitative data also collected
  - Small scale (10s of participants)
- Instrumented panels

# What other options are there?

---

- Field / lab studies
  - 1:1 interaction between researcher and participant
  - Participant performs tasks while researcher watches
  - Observational data (akin to logs) are collected
  - Additional qualitative data also collected
  - Small scale (10s of participants)
- Instrumented panels
  - Users agree to be monitored via some application over some period of time
  - During that period of time, users complete tasks and potentially also answer questions
  - Questions and tasks often determined a priori
  - Medium scale (100s to 1000s of participants)

# Many advantages to these studies:

---

- Labeled data!
  - Specific intent statements
  - Satisfaction / happiness
- Additional quantitative and qualitative data
  - Preferred search engines
  - Self-labelled search expertise
- Information not restricted to the search engine web site
- Can understand specific behavior (why that query?)
  
- Field/lab studies and instrumented panels are much better at getting data for actually inferring user intent!

# But...

---

- While better for inferring intent, there are other issues:
  - Choose stratification a priori
  - User sample is highly biased
  - Choosing tasks vs. “own” tasks
  - Specific task boundaries (missing multi-tasking behavior?)
  - A lot of manual effort

# Conclusion

---

- Query logs still necessary!
  - Immediacy / timeliness
  - Diversity / scalability
  - Meaning disambiguation
  - Aggregation: key component to mitigating privacy concerns
- BUT: more data needed to augment query logs
  - Alternative data sources needed, especially for intent
  - Further research needed in inferring intent, likely should not depend on query logs (insufficient information)
- Ultimately: need multiple data sources!
  - Use field studies / panels to generate hypotheses
  - Can determine prevalence of hypotheses in the logs