

The background features a glowing blue world map with a white grid of latitude and longitude lines. In the foreground, the silhouettes of six business professionals are shown against the map. From left to right: a man in a suit, a woman in a dress, a woman in a suit shaking hands with a man in a suit, a woman in a dress, and a woman in a suit holding a briefcase. The overall color scheme is a gradient of blues.

# Introduction Madness

World Wide Web 2007  
Workshop on Query Log Analysis

# Method to the Madness

- Goal: To quickly get to know each other
- One minute maximum
- I will warn you at 55 seconds
- No animation
- Have fun!



# Order of Introductions

- Jaime Teevan
  - G. Craig Murray
  - Yabo-Arber, Xu
  - Min Zhang
  - Yiqun Liu
  - Wei Vivian Zhang
  - Susan Dumais
  - Jalal Mahmud
  - Li Xiong
  - Michael Maslov
  - Diane Tang
  - Carrie Grimes
  - Daniel Russell
  - Rosie Jones
  - Ryen White
  - Martin Arlit
- 
- The background of the slide features a light blue world map with a grid of latitude and longitude lines. In the foreground, there are several dark silhouettes of people. On the left, a man stands with his hands in his pockets. In the center, a woman in a business suit is shaking hands with a man. To the right, a group of people is standing in a line, with one person holding a briefcase. The overall scene suggests a professional or academic gathering.

# Jaime Teevan

Microsoft Research

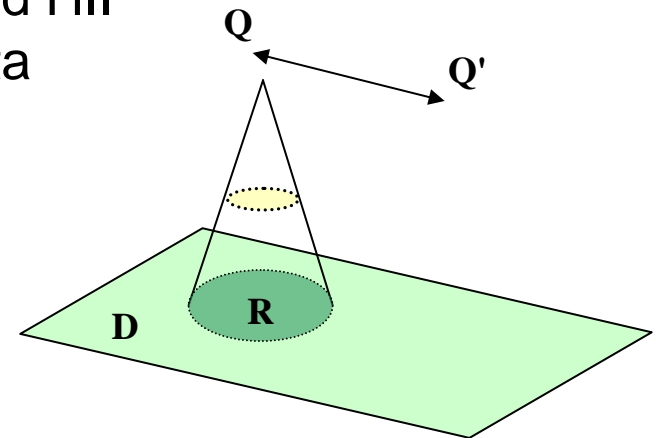
- Research interests:
  - Finding and re-finding
  - Personal information management
- Log analysis important part of my research
- People seem generally happy to give up some privacy for personal benefit
- Sharing probably mostly a matter of policy
- Challenge: Short-term, workable solution



# G. Craig Murray

Student Research Fellow  
University of Maryland

- PhD Student, College of Information Studies
- General research interest
  - Human-Information Interaction
  - Machine learning techniques for improved HII
  - Gathering implicit and explicit human data for modeling relevance
- Interest in Query Logs
  - *Research:*
    - modeling query behavior
    - predicting query trajectories
  - *Policy:*
    - fair and safe practices for happy users
    - workshops to bring together people like you



## **P**ERSONAL PROFILE

---

Name

**Yabo-Arber, Xu**  
**4th-year doctoral student**

General Interests

**Preserving individual privacy in the context of information sharing and exchange**

Specific Interests on QL Analysis

**Anonymizing QLs ( from the service side);**  
**Blocking QL recording( from the client side)**

Work to be presented on WWW07

**Privacy-Enhancing Personalized Web Search**

# Judit Bar-Ilan, Bar-Ilan University

- Background: in CS, now Information Science
- Interest is query logs: information behavior
- Why I'm here? To learn about your views and to share mine
- Research on QL: Website analysis (SHIL)
- How to share QL? Within a well defined framework, respecting the privacy of individual users
- Why to share QL? To allow meaningful academic research
- Why NOT to share? Privacy statements, commercial advantage



# Min Zhang

Ph.D and assistant professor from Tsinghua University, Beijing, China.

[z-m@tsinghua.edu.cn](mailto:z-m@tsinghua.edu.cn)

## Recent work:

- Using query log and click-through data analysis to:
  - identify search engine users' information need types
  - evaluate search engine performance automatically
  - separate key resource pages from others
  - estimate Web page quality

## Our Lab:

- A joint lab
- R&D Support to a widely-used Chinese Search Engine Sogou.com



北京2008年奥运会赞助商  
OFFICIAL SPONSOR OF THE BEIJING 2008 OLYMPIC GAMES



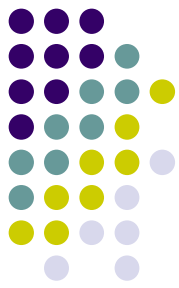
# Yiqun Liu

Ph.D student from Tsinghua University, Beijing, China.

[liuyiqun03@gmail.com](mailto:liuyiqun03@gmail.com)

- **Query and click-through logs got from Sogou.com**
- **Query type identification**
  - Identify the type of user's information need
  - Extract user behavior features from logs
  - Over 80% queries are correctly classified
- **Search engine performance evaluation**
  - A cranfield-like evaluation approach
  - Construct query topic set and answer set Automatically .
  - Obtain similar evaluation results with manual based methods, and cost far less time and labor.

# Wei Vivian Zhang



## Yahoo!

- Research interests using query logs
  - Query Rewriting
  - Session Analysis
  - Geo-query Analysis
- Present “Comparing Click Logs and Editorial Labels for Training Query Rewriting” today

# Susan Dumais

Microsoft Research

<http://research.microsoft.com/~sdumais>



- Research interests
  - Algorithms and interfaces for information retrieval (desktop, enterprise and web)
- Logs ...
  - Compliment other techniques for understanding and improving info access
  - Represent the wide diversity of user behavior *in situ*
- Recent research using logs
  - M. Sanderson and S. Dumais (2007). [Examining repetition in user search behavior.](#) *ECIR 2007*.
  - D. Downey, S. Dumais and E. Horvitz (2007). [Models of searching and browsing: Languages, studies and application.](#) *IJCAI 2007*, 2740-2747.
  - E. Agichtein, E. Brill, S. T. Dumais and R. Ragno (2006). [Learning user interaction models for predicting web search preferences.](#) *SIGIR 2006*, 3-10.
  - E. Agichtein, E. Brill and S. T. Dumais (2006). [Improving web search ranking by incorporating user behavior.](#) *SIGIR 2006*, 19-26.
  - F. Radlinksi and S. T. Dumais (2006). [Improving personalized web search using results diversification.](#) *SIGIR 2006*, 691-692.
  - E. Cutrell, D. C. Robbins, S. T. Dumais and R. Sarin (2006). [Fast, flexible filtering with Phlat - Personal search and organization made easy.](#) *CHI 2006*.
  - J. Teevan, S. T. Dumais and E. Horvitz (2005). [Beyond the commons: On the value of personalizing Web search.](#) *Workshop on New Technologies for Personalized Info Access*.
  - S. Fox, K. Karnawat, M. Mydland, S. T. Dumais and T. White (2005). [Evaluating implicit measures to improve the search experience.](#) *ACM:TOIS*, 23(2), 147-168.
  - X. Shen, S. T. Dumais and E. Horvitz (2005). [Analysis of topic dynamics in web search.](#) *WWW 2005*.



**Jalal Mahmud**

**PhD Student**

Department of Computer Science

Stony Brook University

Stony Brook, NY, USA

Web: [www.cs.sunysb.edu/~jmahmud](http://www.cs.sunysb.edu/~jmahmud)

Email: [jalal.mahmud@gmail.com](mailto:jalal.mahmud@gmail.com)


Antonio Gulli ([agulli@ask.com](mailto:agulli@ask.com))

- Director, Advanced Search Projects, Ask.com
- Head of Pisa European R&D center



# Eytan Adar

Doctoral Student, University of Washington

- Previously at HP Labs & Xerox PARC
  - Interested in:
    - Large group behavior on the Internet
    - IR & Data Mining
    - Social Networks
    - Privacy, Visualization, P2P, ...
- 

# Li Xiong

Assistant Professor, Math & CS Department, Emory University

- *General research interests:* information and data management, privacy, trust, medical and bioinformatics
- *Query log interests:* privacy preserving query log publishing (in collaboration with query log researchers)
- *Workshop:* learn about query log analysis, glaciers, mountains and lakes!
- *Position:* Query logs should be shared but should be anonymized (listen to my presentation!)

# Michael Maslov

Yandex, <http://www.yandex.ru>

- What is Yandex?
  - #1 search engine in Russia (Google is #3)
  - 45 mln SERP hits daily
  - 5 mln users daily, mainly Russian-speaking
- Our goals of query log analysis:
  - search quality
  - ads effectiveness
  - online news processing: [WWW2006: Extracting News-Related Queries from Web Query Log](#)
- How we share query logs:
  - Yandex “Internet Mathematics” grants program, 30+ grants yearly
  - “search Engine Logs” dataset for scholars:
    - sample: ~50 mln queries along the some week, all queries from some part of users
    - data is grouped hierarhically: users -> queries -> url\_shows&clicks
  - scholars ought to sign Yandex NDA

# *Are query logs enough?*



**Diane Tang:** ads-quality team, logging infrastructure, how to analyze session behavior, creating metrics & models of user behavior



**Carrie Grimes:** search quality statistical insights & analysis,



**Daniel M. Russell:** search quality from the user's perspective, field studies, lab studies, models of user behavior





# Rosie Jones, Yahoo! Research

- Background: Information Retrieval, Machine Learning, Computational Linguistics
- Characterizing query reformulation, word deletion (SIGIR 2003)
- Query similarity from reformulation (WWW 2006)
- Query similarity from time series (CIKM 2006)
- Geo-information in queries (GIR 2006 workshop, IGJIS 2007)
- Repeated searches and clicks for personalization (SIGIR 2007, with Jaime Teevan, Eytan Adar and Michael Potts)
- Today: Click data as training signal for machine learning (query log workshop 2007)
- Ongoing: Privacy issues in query logs

# Ryen White – Microsoft Research

- Researcher in Text Mining, Search, and Navigation Group at MSR Redmond
- Interested in:
  - Surfacing information about many users' behavior to benefit the individual
    - *“10% of people who searched for X also did Y”*
- At the workshop to discuss:
  - User privacy
  - Alternatives to raw logs

# Martin Arlitt

University of Calgary and HP Labs

- Interests are workload characterization and performance/scalability evaluation. In particular, I've characterized web servers, web proxies, e-commerce servers, etc.
- I have not done much in terms of query analysis, but I think there is obvious overlap.
- I have shared logs via:
  - the Internet Traffic Archive (<http://ita.ee.lbl.gov>)
  - W3C WCA Workload Characterization Repository (<http://www.cs.vt.edu/repository> - may now be defunct).
- In the network measurement community there is the "Internet Measurement Data Catalog"
  - <http://www.datcat.org/>

The background of the slide features a glowing blue world map with a white grid of latitude and longitude lines. In the foreground, there are black silhouettes of six business professionals. On the far left, a man in a suit stands with his hands in his pockets. To his right, a woman in a dress stands with her back to the camera. In the center, a woman in a business suit is shaking hands with a man in a suit. To the right of the handshake, another woman in a business suit stands with her back to the camera. On the far right, a woman in a business suit stands with her back to the camera, holding a briefcase. The overall scene suggests a global business meeting or conference.

# Introduction Madness

World Wide Web 2007

Workshop on Query Log Analysis