

Investigating the Querying and Browsing Behavior of Advanced Search Engine Users

Ryen W. White, Dan Morris
Microsoft Research, Redmond, USA
{ryenw,dan}@microsoft.com



Motivation

- Some people are more expert at searching than others
 - Search expertise, not domain expertise
- We study characteristics of these “**advanced search engine users**” in an effort to better understand how these users search
- If we can better understand what advanced searchers are doing maybe we can improve the search experience for everyone

Study

- We define **advanced users** as searchers who use advanced query operators in their query statements,
 - Used plus (+), minus (-), quotes (“”), and “site:”
 - e.g., “sigir 2007”
 - e.g., microsoft +office site:microsoft.com
- Is there a relationship between the use of advanced syntax and:
 - Queries and result clicks?
 - Post-query browsing?
 - Search success?

Data

- **Interaction logs** of 586K opt-in users
 - English U.S. users
 - 13-week period from January to April 2006
 - Complete browsing history
 - Search engine queries (Multiple engines)
 - All page visits
- **Relevance judgments** for 11K queries
 - 6-level judgments

Characterizing Advanced Users

- Four advanced operators used: +, -, “”, and “site:”
- ~1% of submitted queries contained at least one operator
- 51K users (9%) of users used query operators at least once
- p_{advanced} used to denote the percentage of a user's queries that contain advanced operators
 - Non-advanced users ($p_{\text{advanced}} = 0\%$)
 - Advanced users ($p_{\text{advanced}} > 0\%$)
- Included users who issued > 50 queries
 - ~38K (20%) advanced users
 - ~151K (80%) non-advanced users



Do advanced users query and click on results differently than novices?

Query and Result-Click Features

- Give overview of subjects' direct interactions with search engines

Feature	Meaning
Query Repeat Rate (QRR)	Fraction of queries that are repeats
Query Word Length (QWL)	Avg. number of words in query
Queries Per Day (QPD)	Avg. number of queries per day
Queries Per Second (QPS)	Avg. number of queries per second between initial query and end-of-session
Avg. Click Position (ACP)	Avg. rank of clicked results
Click Probability (CP)	Ratio of result clicks to queries
Avg. Seconds To Click (ASC)	Avg. search to result click interval

Findings: Query/Result-click

Feature	P _{advanced}	
	0%	> 0%
QRR	0.53	0.57
QWL	2.02	2.83
QPD	2.01	3.52
QPS	0.028	0.010
ACP	6.83	9.12
CP	0.57	0.51
ASC	87.71	88.16
%Users	79.90%	20.10%

Non-advanced Advanced

- Advanced users:
 - Repeat queries more often
 - Compose longer queries
 - Submit more queries/day
 - Query less/second
 - Click further down the result list
 - Less likely to click a result

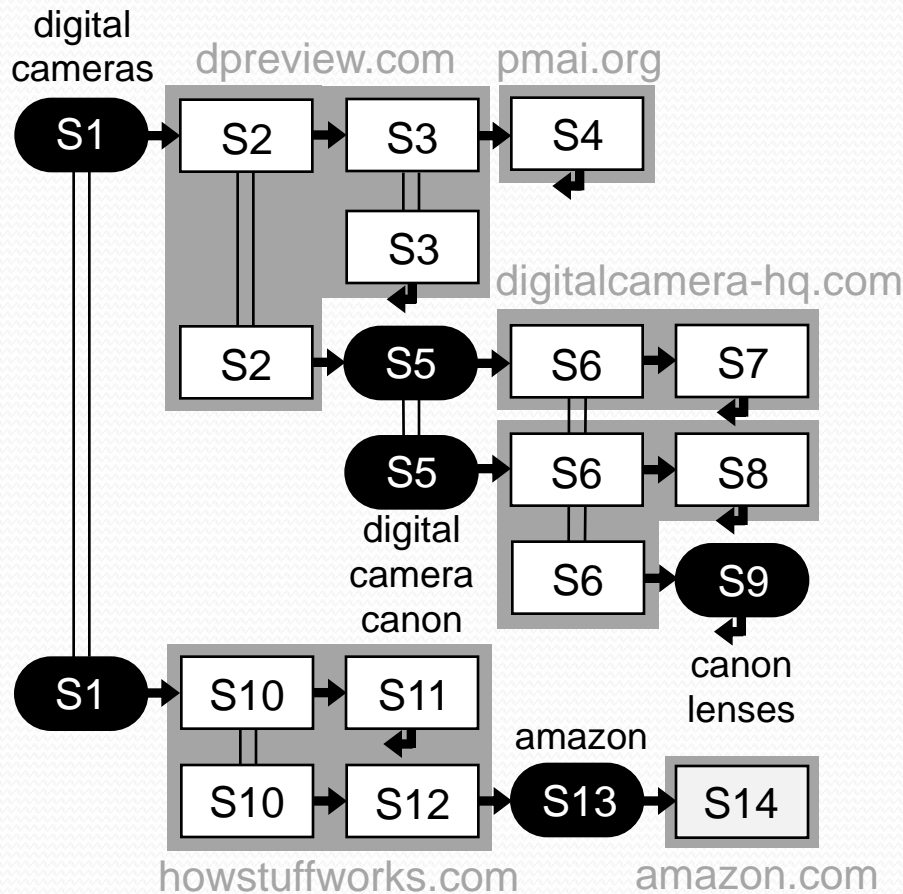
Findings: Query/Result-click

- **Factor analysis** to study the relationships among the dependent variables
- Factor analysis revealed two factors that could account for ~84% of the variance:
 - Factor A = Querying
 - Query properties associated with position of clicks in result list
 - Factor B = Result-click
 - Querying frequency associated with the likelihood that user will click on a search result and click latency



Do advanced users browse
differently than novices?

Search Session



- **Session**
 - Query → Timeout
- **Query trail**
 - Query → End trail event
 - Another query
 - Type URL
 - Visit homepage
 - Check Web-based email or logon to online service
 - Close browser
 - Session timeout

Post-Query Browsing Features

- Based on search sessions and search trails extracted from interaction logs
- 12.5 million search trails extracted
- Median number of trails per user was 30
- Median number of steps in the trails was 3

Post-Query Browsing Features

Feature	Meaning
Session Seconds	Average session length (in seconds)
Trail Seconds	Average query trail length (in seconds)
Display Seconds	Average display time for each page on the trail (in seconds)
Num. Steps	Average number of steps from the page following the results page to the end of the trail
Num. Revisits	Average number of “back” operations
Num. Branches	Average number of branches

Findings – Post-query browsing

Feature	P _{advanced}	
	0%	> 0%
Session Secs	701.10	706.21
Trail Secs	205.39	159.56
Display Secs	36.95	32.94
Num. Steps	4.88	4.72
Num. Revisits	1.20	1.02
Num. Branches	1.55	1.51
% _{Trails}	72.14%	27.86%
% _{Users}	79.90%	20.10%

Non-advanced Advanced

- Advanced users:
 - Traverse trails faster
 - Spend less time viewing each Web page
 - Follow query trails with fewer steps
 - Revisit pages less often
 - “Branch” less often

Findings – Post-query browsing

- Greater the proportion of queries with advanced syntax the more focused their search interactions become
 - Shorter query trails
 - Less “branchy” query trails
- Session time increases but search time drops with increases in p_{advanced}
 - Perhaps more advanced users are multitasking between search and other activities



Are advanced users more
successful than novices?

Search Success

- Human relevance judgment available for 11K queries
 - Extract corresponding query trails from our logs
 - Relevance judgments for 56% of pages on those trails
- We use these judgments to compute several metrics for search success

Metric	Meaning
First	Judgment assigned to the first page in the trail
Last	Judgment assigned to the last page in the trail
Average	Average judgment across all pages in the trail
Maximum	Maximum judgment across all pages in the trail

Findings – Search Success

- Average relevance judgment, Min = 1, Max = 6
- More advanced users are more likely to have success

Feature		P _{advanced}	
		0%	> 0%
First	<u>M</u>	4.03	4.19
	<u>SD</u>	1.58	1.56
Last	<u>M</u>	3.79	3.92
	<u>SD</u>	1.60	1.57
Max	<u>M</u>	4.04	4.20
	<u>SD</u>	1.63	1.51
Avg.	<u>M</u>	3.93	4.06
	<u>SD</u>	1.57	1.51

Non-advanced Advanced

Summary

- **Conducted** log-based study of search behavior
- **Classified** users according to their use of advanced query syntax
- **Demonstrated** that use of advanced search syntax is correlated with other aspects of search behavior such as queries and result clicks, post-query navigation, and search success
- **Next steps**
 - Use the interactions of advanced users for improved document ranking, page recommendation, training

Thank you!

Questions/comments?

Ryen W. White, Dan Morris
{ryenw, dan}@microsoft.com

