

Web Searching: Will I Be Able to Find It on the Web?

Amanda Spink
School of Information
Sciences
University of Pittsburgh

1

Maybe – Maybe Not!

It may depend on YOU!

Search improvement will only come through a combination of systems development and user effort

Needs to be more focus on improving user effort

2

Why Maybe - Maybe Not?

- Reason 1: Web search is hard for both users and Web search engines designers
- Reason 2: No current technical silver bullet to improve Web search
- Reason 3: Web search engines do not inform users well about the Web search basics
- Reason 4: Many people do not Web search effectively, understand Web search basics, and are impatient in their information behaviors

3

How Do We Know This?

Twenty years experience in the electronic search business

Amanda Spink & Bernard J. Jansen (2004).
Web Search: Public Searching of the Web
Published by Springer

Based on many Web search studies from 1997 to 2004
– only research book to look at peoples' Web search behavior

4

My Web Search Research Partners

- Large-scale Web studies from 1997 – 2004
- Excite.com
- AlltheWeb.com
- AskJeeves.com
- AltaVista.com
- Vivisimo.com

5

Research Goals

- Track Web search trends – 1997 to 2004
- Identify characteristics of Web searching – terms, queries, sessions, use of query operators, subjects searched, etc.
- Implications for understanding Web search trends and characteristics on a large-scale.

6

Web Query Datasets

- Web query transaction logs from 1997 - 2004
- Dataset of 30 million+ Web queries from multiple Web search engines
- How are people searching the Web?

7

Queries Per User

Queries	97	99	01
Mean	2.8	2.5	2
1 query	67%	48%	78%
2	19%	21%	13%
3	7%	11%	4%

SHORT SEARCHES: 2004 – Most users enter 1 or 2 queries

8

Terms Per Query Distribution

Terms	97	99	01
Mean	2.32	2.4	2.35
1 term	31%	26%	26%
2	31%	31%	26%
3	18%	18%	15%

SHORT QUERIES: 2004 – 70% users enter 3 or less terms per query - mean 2.5 terms per query

9

Web Search Session Duration - 2004

- 50% sessions less than 1 minute
- 10% sessions 1-5 minutes
- 45% sessions longer than 5 minutes

10

Use of Boolean Operators

	97	99	01
In >10% of queries		>5%	20%

2004 – still low use of Boolean, but higher use of other advanced features

Many uses of Boolean operators are wrong - not according to instructions how to use them

11

Pages Viewed Per User

Pages	97	97	01
Mean	2.3	1.8	1.7
1 page	58%	29%	43%
2	19%	19%	21%
3	9%	14%	

2004 - Most users view VERY FEW pages beyond the first or first two pages.

12

Top 10 Terms By Frequency

97

sex
nude
free
pictures
new
university
women
chat
gay
girls

99

sex
free
nude
pictures
university
pics
chat
adult
women
new

01

sex
Christmas
nude
pictures
new
pics
music
university
games
porn

13

Top 10 Terms By Frequency - 2004

- download
- new
- software
- windows
- sex
- school
- history
- online
- video
- what

14

Query Subjects

<u>Subject Category</u>	<u>97</u>	<u>99</u>
1. Entertainment, recreation	16.9%	7.5% (6)
2. Sex, porn, preferences	16.8%	7.5% (4)
3. Commerce travel, economy	13.3%	24.4% (1)
4. Computers & the Internet	12.5%	10.9% (3)
5. Health & the sciences	9.5%	7.8% (5)
6. People, places, things	6.7%	20.3% (2)
7. Society, culture, religion	5.7%	4.2% (9)
8. Education & the humanities	5.6%	5.3% (8)
9. Performing & fine arts	5.4%	1.1% (11)
10. Government	3.4%	1.6% (10)
11. Incomprehensible	4.1%	6.8% (7)

15

Query Subjects - 2004

• Commerce, Travel, Employment or Economy	21%
• Indiscernible or Non-English	19%
• People, Places or Things	15%
• Computers or Internet	13%
• Social, Culture, Ethnic or Religion	9%
• Health or Sciences	6%
• Education or Humanities	5%
• Sex or Pornography	4%
• Performing or Fine Arts	3%
• Government	3%
• Entertainment or Recreation	2%

16

Major Findings

- Web searches and queries are short
- Users not viewing many pages
 - mean 1.9 pages - percentage of views falling
 - 1 in 2 users did not go beyond the first page
- Boolean queries not used much
 - Less than 1 in 20 queries
- From 1997 to 2004 in many respects Web searching behavior did NOT change much

17

Query Trends 1997-2004

- Shift from entertainment/sex to e-commerce/people queries
- Growth of non-English queries
- Sex/pornography queries about 5%
- More query reformulation
- Less page viewing
- Why low use of advanced search features? Are they the wrong features?

18

New Web Search Features

- Semantic Web - better categorization of Web content
- Personalization
- Clustering
- New algorithms
- New interface features
- Visualization techniques
- New types of Web search engines

19

Will We Be Able to Find it on the Web?

- Computer science and technologist do not have all the answers or capability to improve Web search
- There is no “silver bullet” – search improvement is tough

20

Will We Be Able to Find it on the Web?

- Human information behavior – need for more understanding, training and knowledge by the average person
- People need to put more effort into their own information behaviors!!
- Many Web search features need redesign or rethinking to accommodate human information behaviors.

21

Conclusions

- Web is a marvelous new technology
 - but people are unpredictable in use of any new technology
 - Improvement will only come through a combination of systems development and user effort
 - how are they **really** using the Web? We have a long way to go to improve Web search.

22



Thank You



Questions?

23