Mining query logs to improve web search engines' operations

Salvatore Orlando[+], Raffaele Perego[*], Fabrizio Silvestri[*]
[*]ISTI - CNR, Pisa, Italy
[+]Università Ca' Foscari Venezia, Italy

Query Log Mining
(for friends :-))

Salvatore Orlando[+], Raffaele Perego[*], Fabrizio Silvestri[*]
[*]ISTI - CNR, Pisa, Italy
[+]Università Ca' Foscari Venezia, Italy

About US

Classes will be given in an ordering obtained by a Rotate Right with Carry operation on this ordering :-)

- **Salvatore Orlando** (*orlando@unive.it*):
  - Professor of CS at University Ca' Foscari, Venezia.
  - Research Interests: Data Mining, Web Mining, Parallel Computing
- **Raffaele Perego** (*raffaele.perego@isti.cnr.it*):
  - Senior Researcher at ISTI - CNR, Pisa.
  - Research Interests: Web Search, Data/Web Mining, Parallel Computing
- **Fabrizio Silvestri** (*fabrizio.silvestri@isti.cnr.it*):
  - Researcher at ISTI - CNR, Pisa.
  - Research Interests: Web Search, Web Mining, "Parallel" Computing

About US

- **Fabrizio Silvestri** (*fabrizio.silvestri@isti.cnr.it*):
  - Researcher at ISTI - CNR, Pisa.
  - Research Interests: Web Search, Web Mining, “Parallel” Computing
- **Salvatore Orlando** (*orlando@unive.it*):
  - Professor of CS at University of Venice.
  - Research Interests: Data Mining, Web Mining, Parallel Computing
- **Raffaele Perego** (*raffaele.perego@isti.cnr.it*):
  - Senior Researcher at ISTI - CNR, Pisa.
  - Research Interests: Web Search, Data/Web Mining, Parallel Computing

Course Plan

- Class 1: Query log analysis.
- Class 2: Query-log based techniques for optimizing WSE effectiveness.
- Class 3: Query-log based techniques for optimizing WSE efficiency.
- Class 4: Hands-on session.
- Class 5: Future Research Issues and the Web of Data.

Course Plan

- Class 1: Query log analysis.
- Class 2: Query-log based techniques for optimizing WSE effectiveness.
- Class 3: Query-log based techniques for optimizing WSE efficiency.
- Class 4: Hands-on session.
- Class 5: ~~Future Research Issues and the Web of Data.~~

Course Plan

- Class 1: Query log analysis.
- Class 2: Query-log based techniques for optimizing WSE effectiveness.
- Class 3: Query-log based techniques for optimizing WSE efficiency.
- Class 4: Hands-on session.
- Class 5: Recent results on the previous topics.

Query log analysis

(Fabrizio Silvestri)

- The first lecture shows the nature of queries submitted by users.
- In particular, it shows how interactions with search engines are done by users in the form of search sessions.

Query-log based techniques for optimizing WSE effectiveness

(Salvatore Orlando)

- query expansion.
- query suggestion.
- results personalization.
- learning to rank.

Query-log based techniques for optimizing WSE efficiency

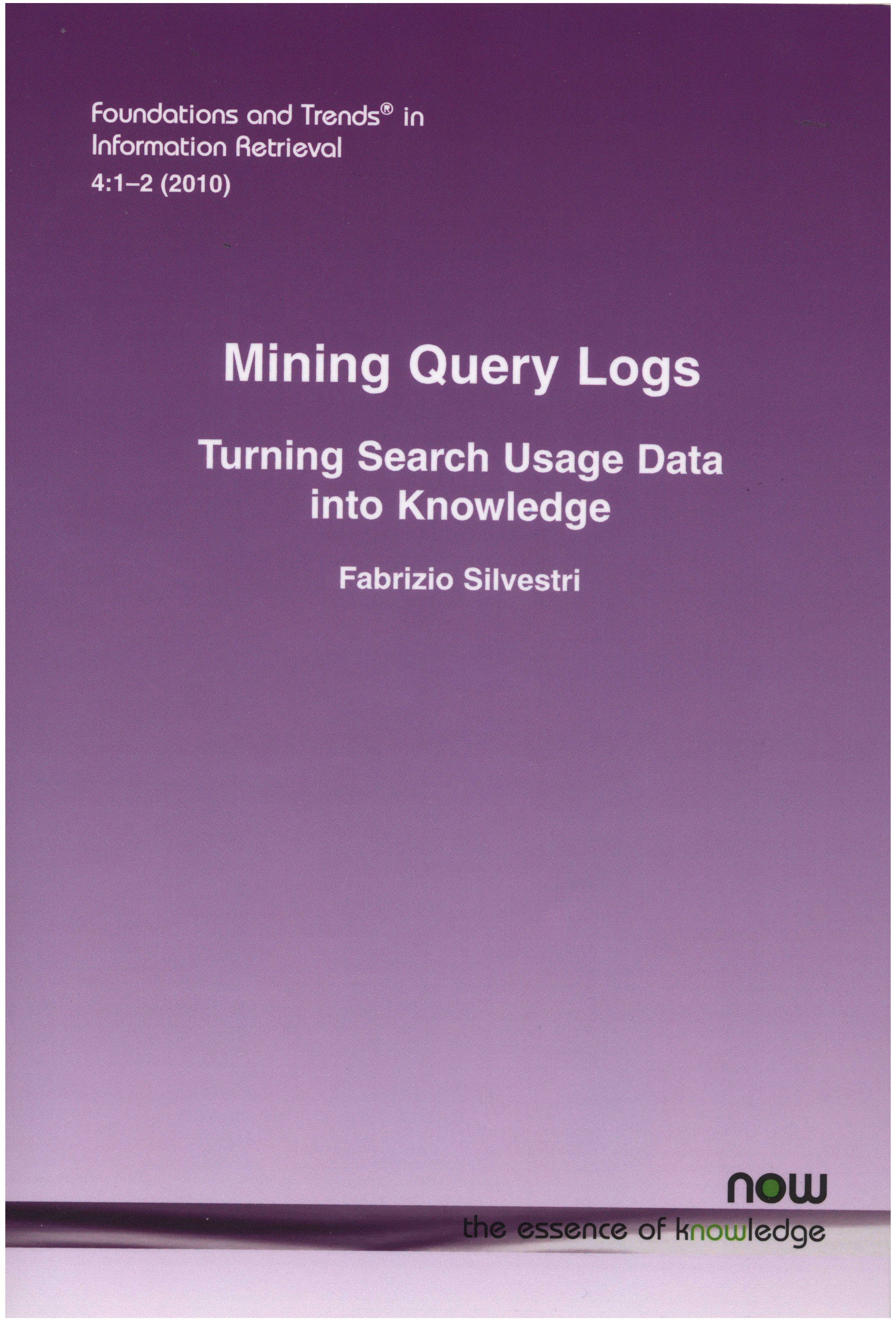
(Raffaele Perego)

- caching in search engines.
- collection partitioning and selection.

Hands-on session

Recent results on Query Log Mining

- We show some novel results and open problems in the field of query log mining
  - possible interesting research directions involve the integration of query log mining and semantic web data analysis research.



- Most of the material is covered by this Book:
    - Fabrizio Silvestri: **Mining Query Logs: Turning Search Usage Data into Knowledge**. Foundations and Trends in Information Retrieval 4(1-2): 1-174 (2010).
- Other relevant papers will be distributed during classes.

Some slides might have been changed/added/removed w.r.t. the ones you have in your handouts!

Questions?

Fasten Your Seat Belts!!!

Query Log Analysis

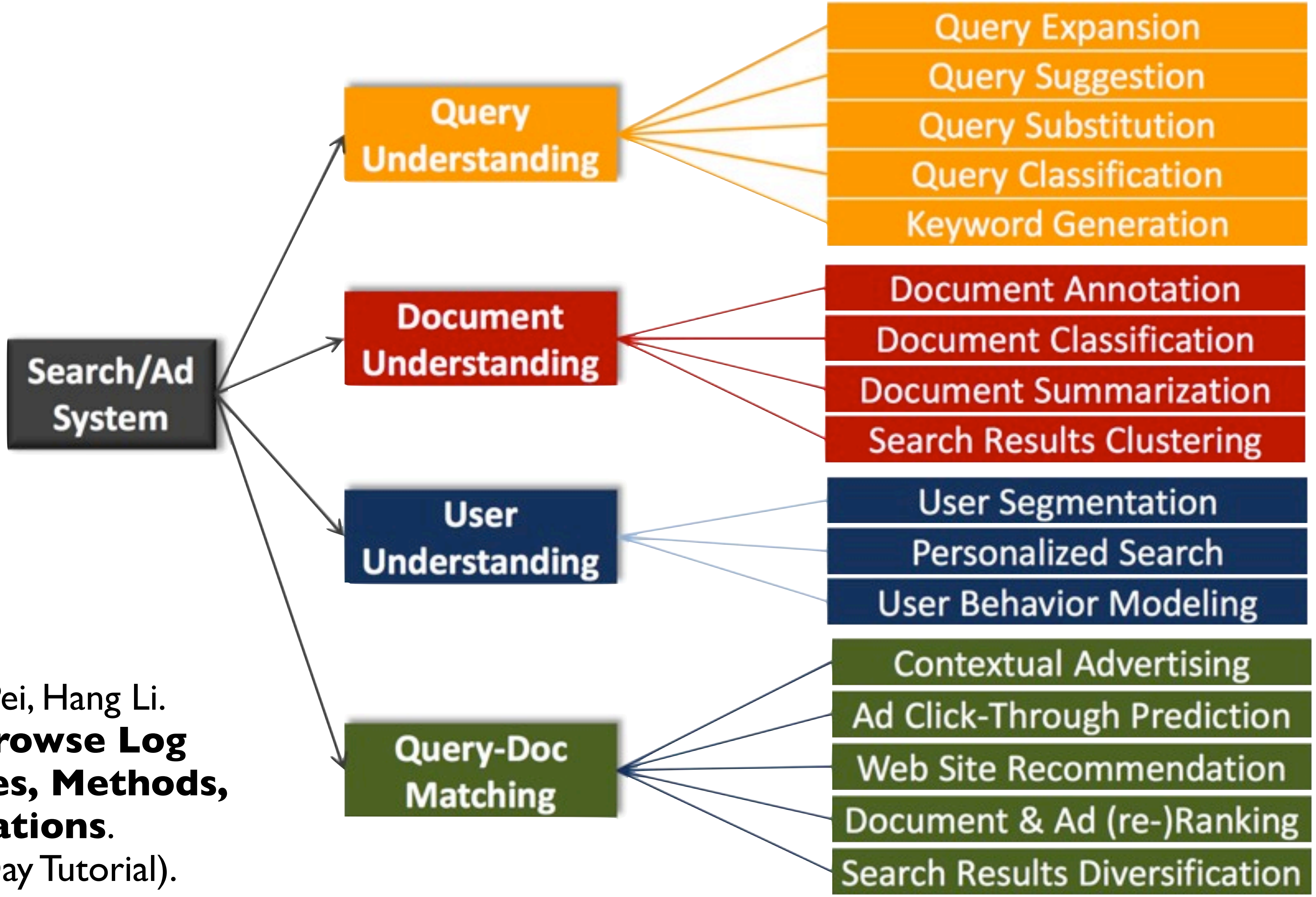
Salvatore Orlando+, Raffaele Perego*, Fabrizio Silvestri*
*ISTI - CNR, Pisa, Italy
+Università Ca' Foscari Venezia, Italy

Web Mining

- Content:
  - text & multimedia mining
- Structure:
  - link analysis, graph mining
- Usage:
  - log analysis, query mining
- Relate all of the above
  - Web characterization
  - Particular applications

Dynamic

Log (Usage) Mining Apps



From:
Daxin Jiang, Jian Pei, Hang Li.
Web Search/Browse Log Mining: Challenges, Methods, and Applications.
WWW'10 (Full-Day Tutorial).

History in Search Engines

Alphonse de Lamartine

Source: Wikipedia

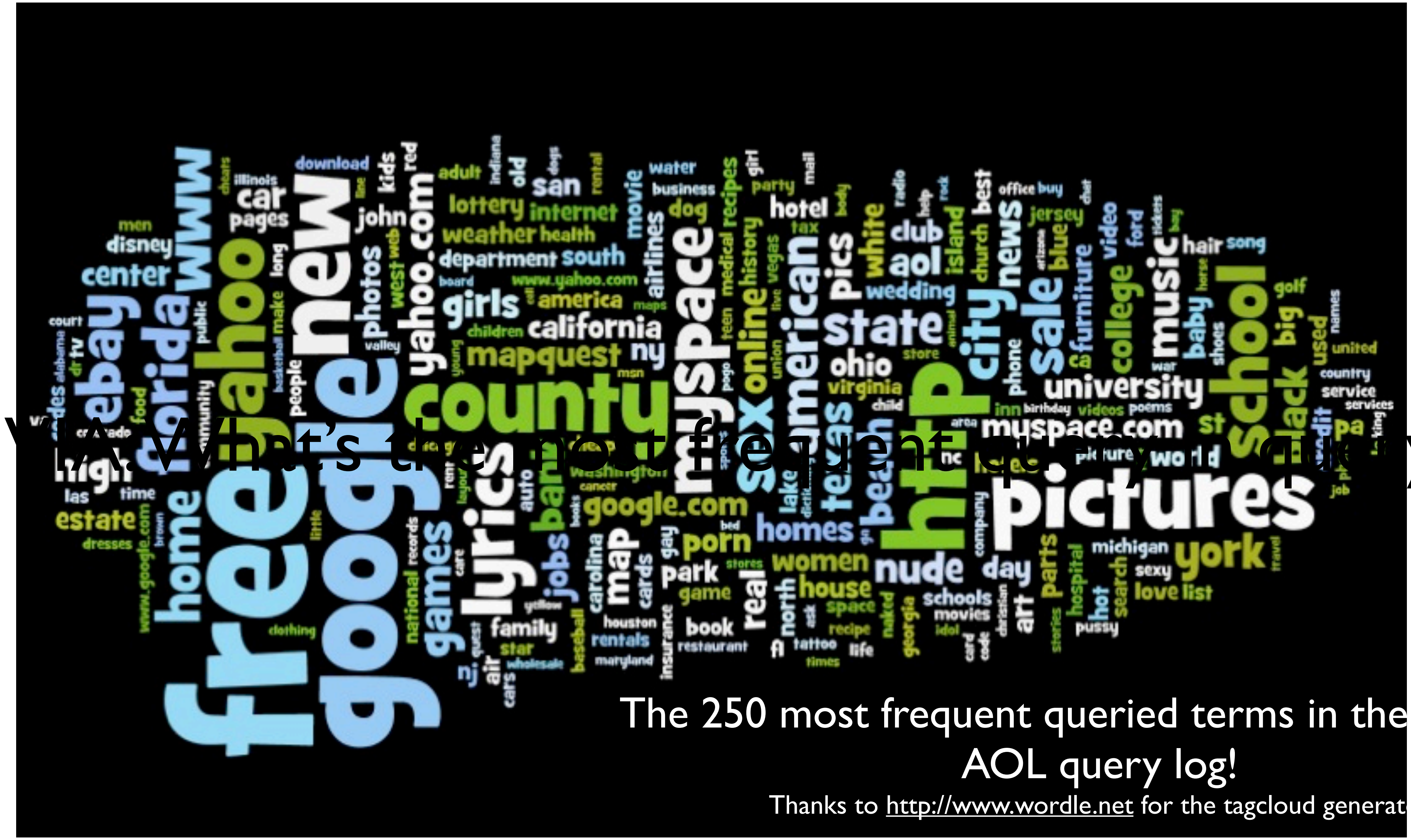
History Teaches Everything... Even the Future!

What is History?

- Past Queries
- Query Sessions
- Clickthrough Data

What’s in Query Logs?

- TRIVIA: What’s the most frequent query in query logs?



The 250 most frequent queried terms in the AOL query log!

Thanks to http://www.wordle.net for the tagcloud generat

Some Examples!

- AOL's us
  - revenge t
  - the woma
  - dirty trick
  - ...
  - locatecell
  - what can
  - mean rev
  - death rec

Some Examples

- AOL User 23187425 typed the following queries within a 10 minutes time-span:
- **you come forward** 2006-05-07 03:05:19
- **start to stay off** 2006-05-07 03:06:04
- **i have had trouble** 2006-05-07 03:06:41
- **time to move on** 2006-05-07 03:07:16
- **all over with** 2006-05-07 03:07:59
- **joe stop that** 2006-05-07 03:08:36
- **i can move on** 2006-05-07 03:09:32
- **give you my time in person** 2006-05-07 03:10:07
- **never find a gain** 2006-05-07 03:10:47
- **i want change** 2006-05-07 03:11:15
- **know who iam** 2006-05-07 03:11:55
- **curse have been broken** 2006-05-07 03:12:30
- **told shawn lawn mow burn up** 2006-05-07 03:13:50
- **burn up** 2006-05-07 03:14:14
- **was his i deal** 2006-05-07 03:15:13
- **i would have told him** 2006-05-07 03:15:46
- **to kill him too** 2006-05-07 03:16:18

I Love Alaska!

- http://www.minimovies.org/documentaires/view/ilovealaska
- “I love Alaska tells the story of one of those AOL users. We get to know a religious middle-aged woman from Houston, Texas, who spends her days at home behind her TV and computer. Her unique style of phrasing combined with her putting her ideas, convictions and obsessions into AOL's search engine, turn her personal story into a disconcerting novel of sorts.

  Over a period of three months, a portrait of a woman emerges who is diligently searching for likeminded souls. The list of her search queries read aloud by a voice-over reads like a revealing character study of a somewhat obese middle-aged lady in her menopause, who is looking for a way to rejuvenate her sex life. In the end, when she cheats on her husband with a man she met online, her life seems to crumble around her. She regrets her deceit, admits to her Internet addiction and dreams of a new life in Alaska.”

Query Logs Analyzed in the Literature

Query log name	Public	Period	# Queries	# Sessions	# Users
Excite ‘97	Y	Sep ‘97	1,025,908	211,063	$\sim 410,360$
Excite ‘97 (small)	Y	Sep ‘97	51,473	N.D.	$\sim 18,113$
Altavista	N	Aug 2^{nd} - Sep 13^{th} ‘98	993,208,159	285,474,117	N.D.
Excite ‘99	Y	Dec ‘99	1,025,910	325,711	$\sim 540,000$
Excite ‘01	Y	May ‘01	1,025,910	262,025	$\sim 446,000$
Altavista (public)	Y	Sep ‘01	7,175,648	N.D.	N.D.
Tiscali	N	Apr ‘02	3,278,211	N.D.	N.D.
TodoBR	Y	Jan - Oct ‘03	22,589,568	N.D.	N.D.
TodoCL	N	May – Nov ‘03	N.D.	N.D.	N.D.
AOL (big)	N	Dec 26^{th} ‘03 – Jan 1^{st} ‘04	$\sim 100,000,000$	N.D.	$\sim 50,000,000$
Yahoo!	N	Nov ‘05 – Nov ‘06	N.D.	N.D.	N.D.
AOL (small)	Y	Mar 1^{st} - May 31^{st} ‘06	36,389,567	N.D.	N.D.

Some Popular Terms: Excite and Altavista

query	freq.
Empty Query	2,586
sex	229
chat	58
lucky number generator	56
p****	55
porno	55
b****y	55
nude beaches	52
playboy	46
bondage	46
porn	45
rain forest restaurant	40
f****ing	40
crossdressing	39
crystal methamphetamine	36
consumer reports	35
xxx	34
nude tanya harding	33
music	33
sneaker stories	32

(a) Excite.

query	freq.
christmas photos	31,554
lyrics	15,818
cracks	12,670
google	12,210
gay	10,945
harry potter	7,933
wallpapers	7,848
pornografia	6,893
"yahoo com"	6,753
juegos	6,559
lingerie	6,078
symbios logic 53c400a	5,701
letras de canciones	5,518
humor	5,400
pictures	5,293
preteen	5,137
hypnosis	4,556
cpc view registration key	4,553
sex stories	4,521
cd cover	4,267

(b) Altavista.

Fabrizio Silvestri: **Mining Query Logs: Turning Search Usage Data into Knowledge**.
Foundations and Trends in Information Retrieval. (To Appear).

Topic Distribution: Excite and AOL

Topic	Percentage
Entertainment or recreation	19.9%
Sex and pornography	16.8%
Commerce, travel, employment, or economy	13.3%
Computers or Internet	12.5%
Health or sciences	9.5%
People, places, or things	6.7%
Society, culture, ethnicity, or religion	5.7%
Education or humanities	5.6%
Performing or fine arts	5.4%
Non-English or unknown	4.1%
Government	3.4%

Excite

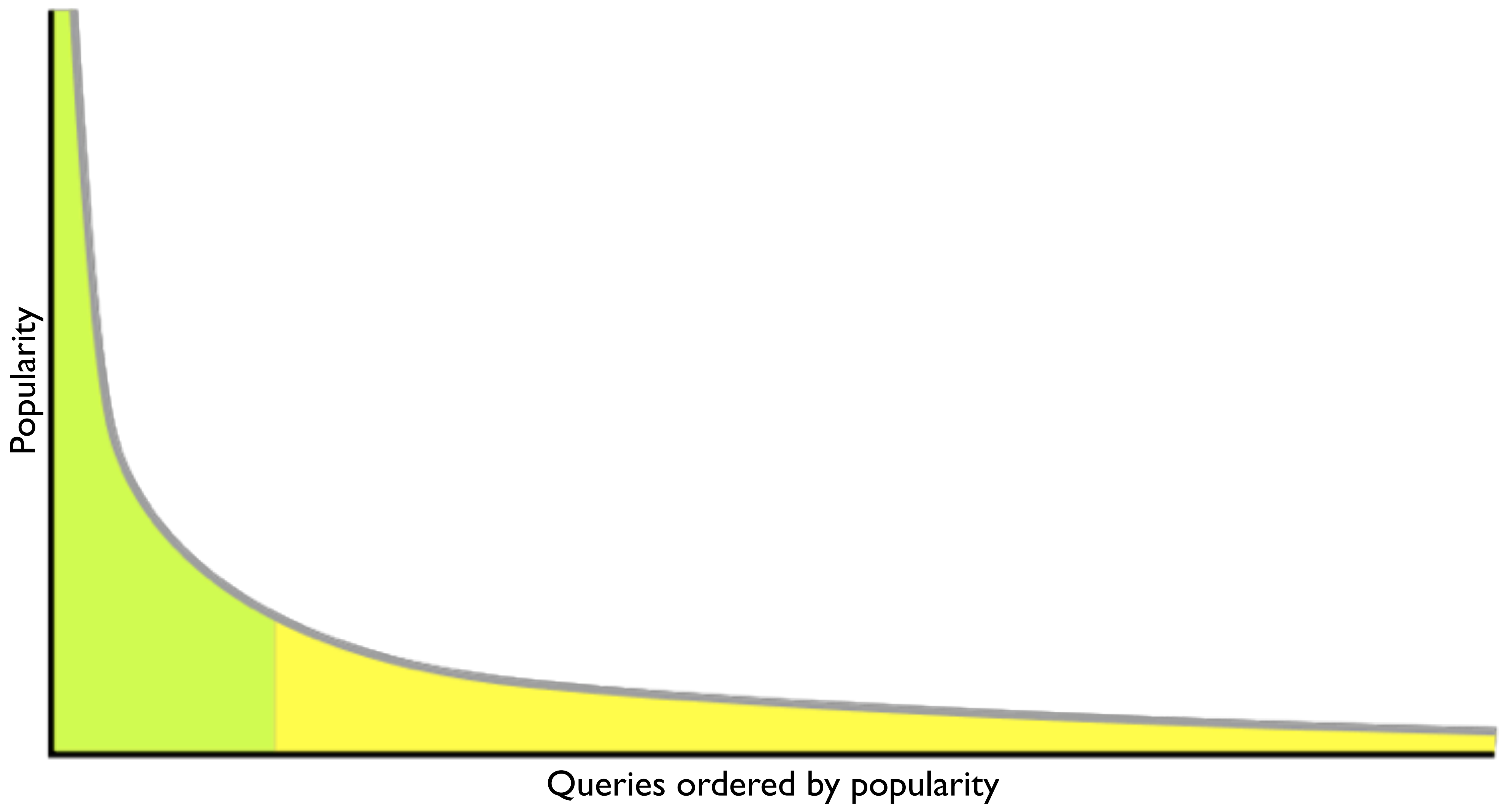
Topic	Percentage
Entertainment	13%
Shopping	13%
Porn	10%
Research & learn	9%
Computing	9%
Health	5%
Home	5%
Travel	5%
Games	5%
Personal & Finance	3%
Sports	3%
US Sites	3%
Holidays	1%
Other	16%

AOL

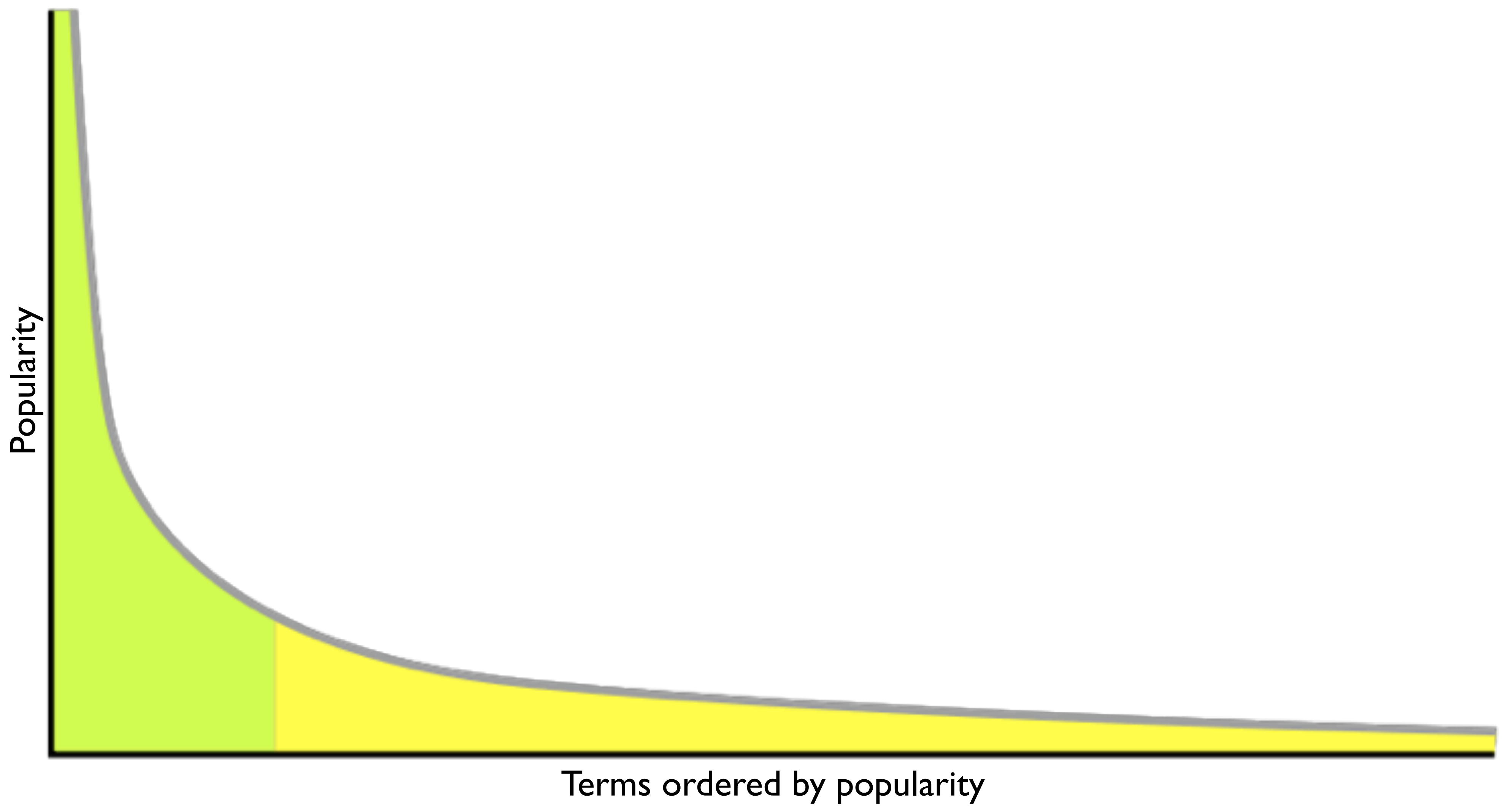
A. Spink, B. J. Jansen, D. Wolfram, and T. Saracevic, "**From e-sex to e-commerce: Web search changes**," Computer, vol. 35, no. 3, pp. 107–109, 2002.

S. M. Beitzel, E. C. Jensen, A. Chowdhury, O. Frieder, and D. Grossman, "T**emporal analysis of a very large topically categorized web query log**," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 2, pp. 166–178, 2007.

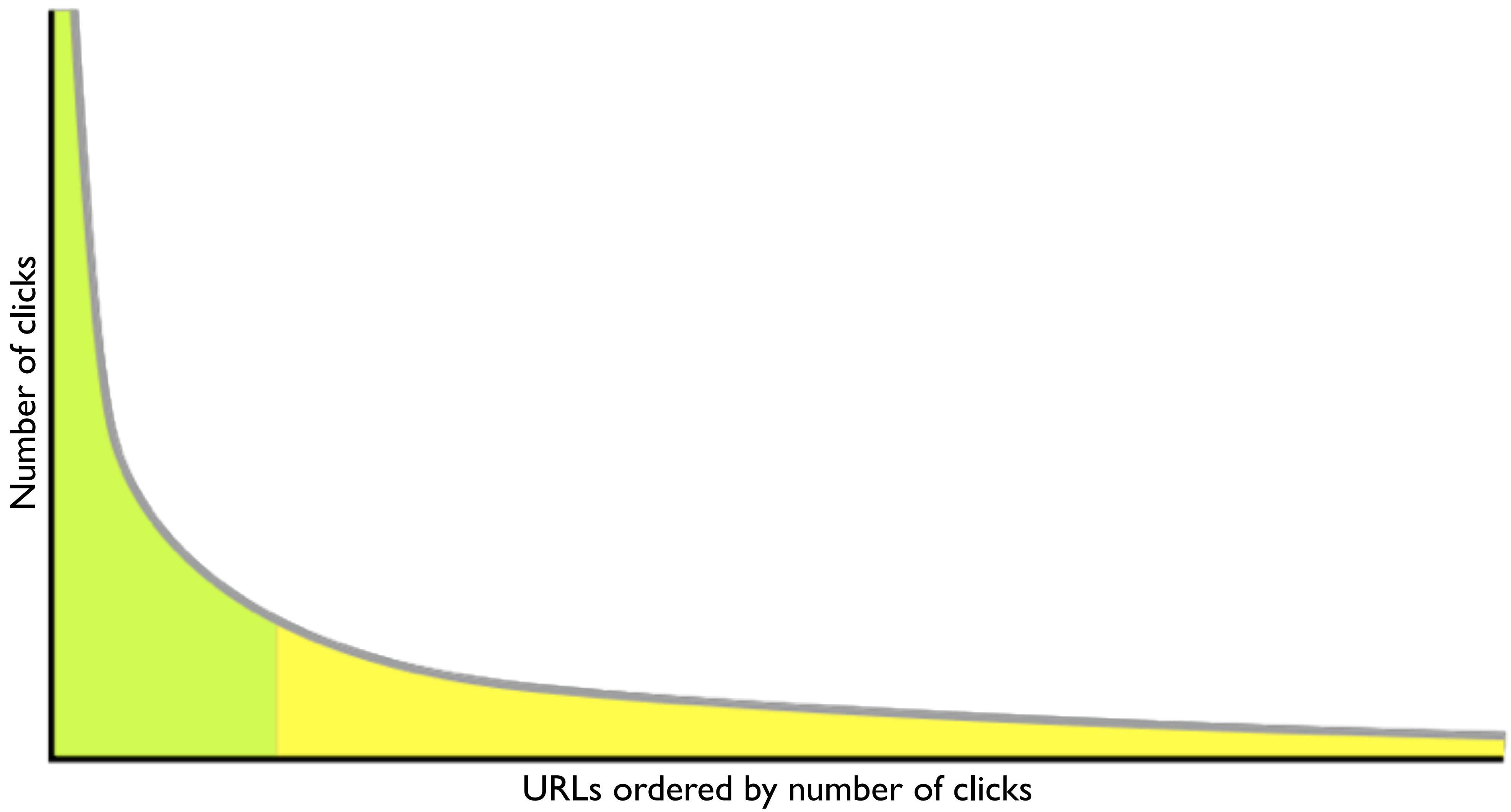
Long Tail Distribution



Long Tail Distribution



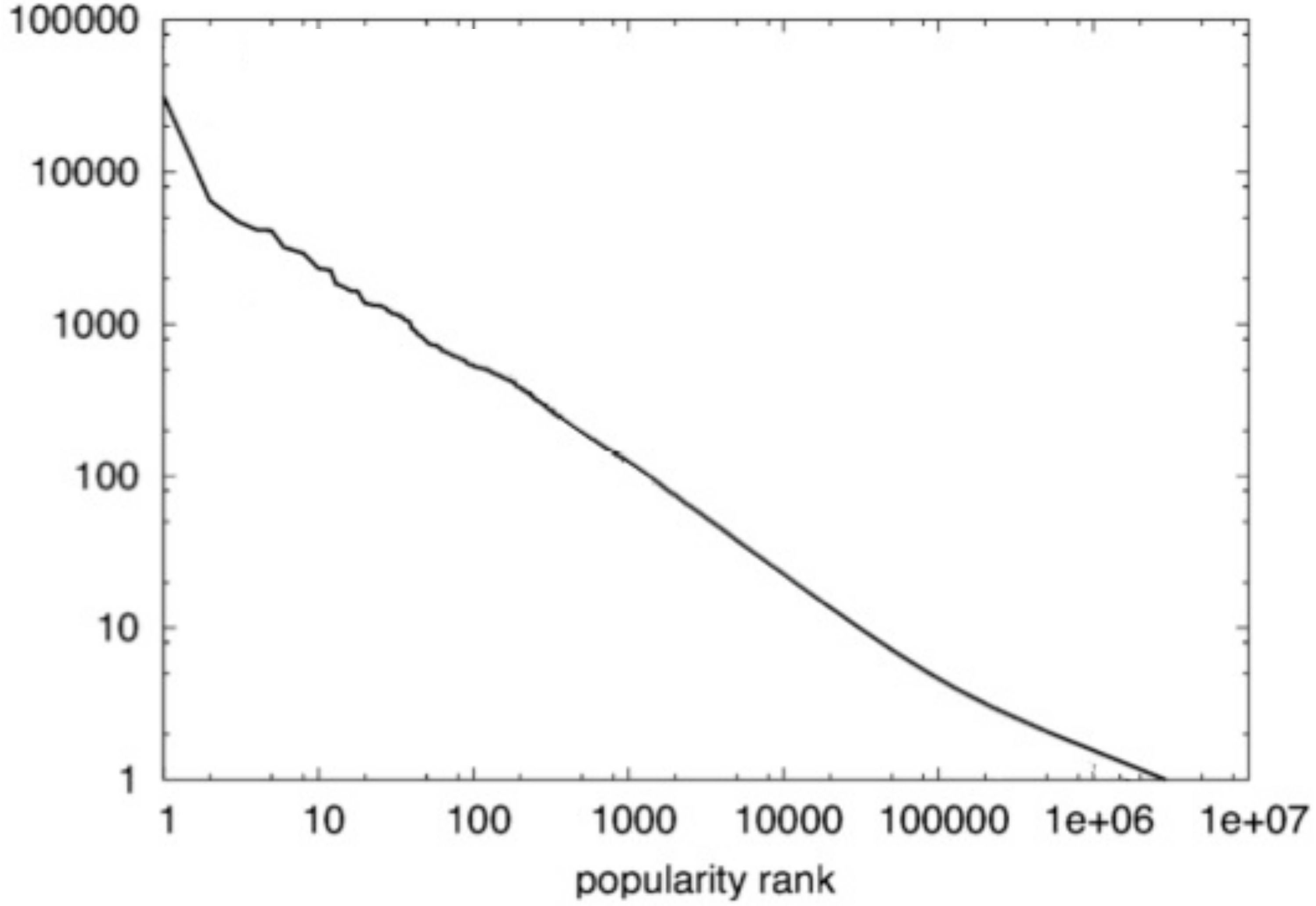
Long Tail Distribution



Power-Laws

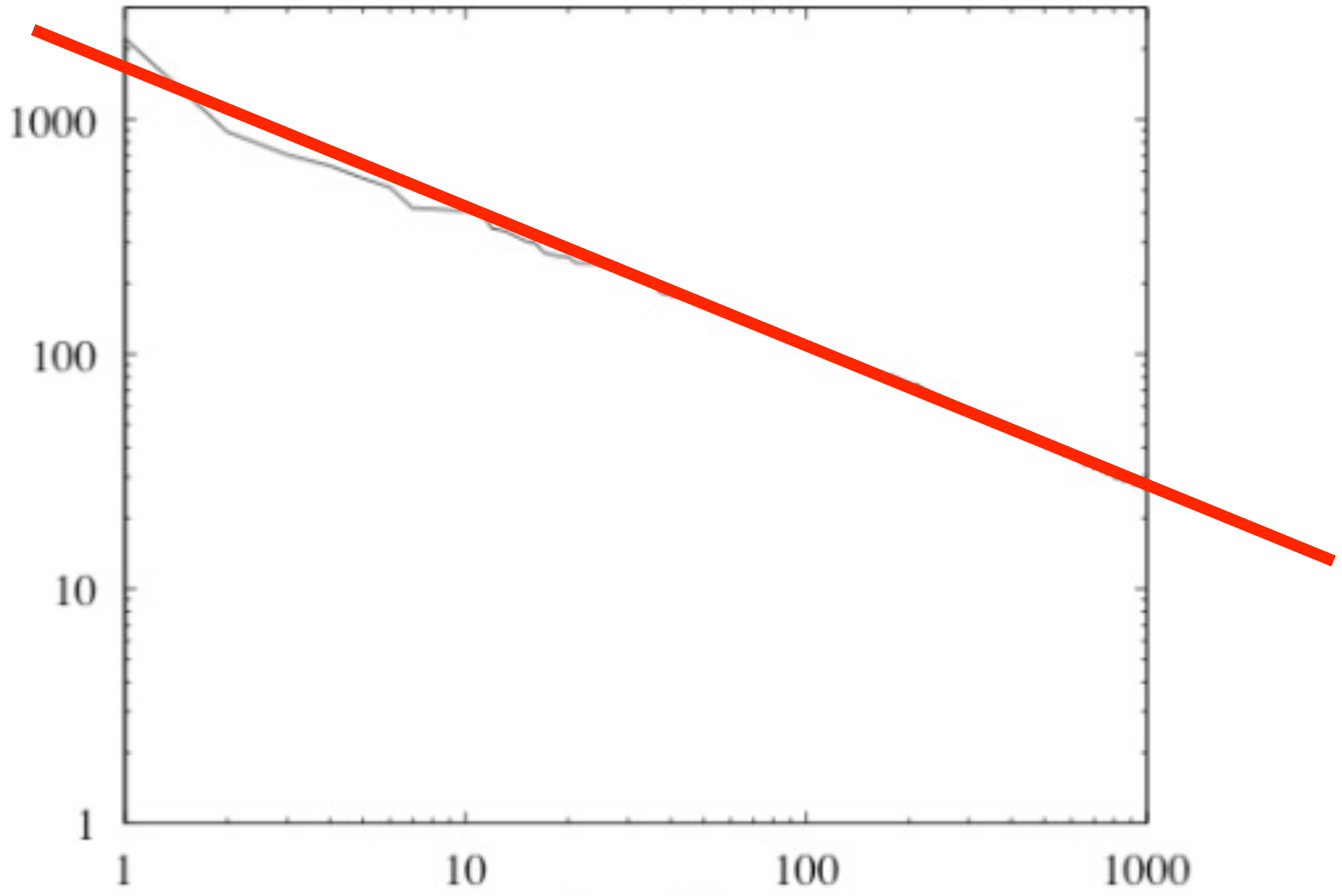
- "When the frequency of an event varies as a power of some attribute of that event (e.g. its size), the frequency is said to follow a power law."
  - Wikipedia's Definition of Power Law
- In practice a D.R.V. X follows a power law if the distribution of X is given by:
  - $P(\{X=x\}) \sim x^{-a}$
  - Exponent "*a*" is the power-law parameter

Power-Law In Query Popularity: Altavista



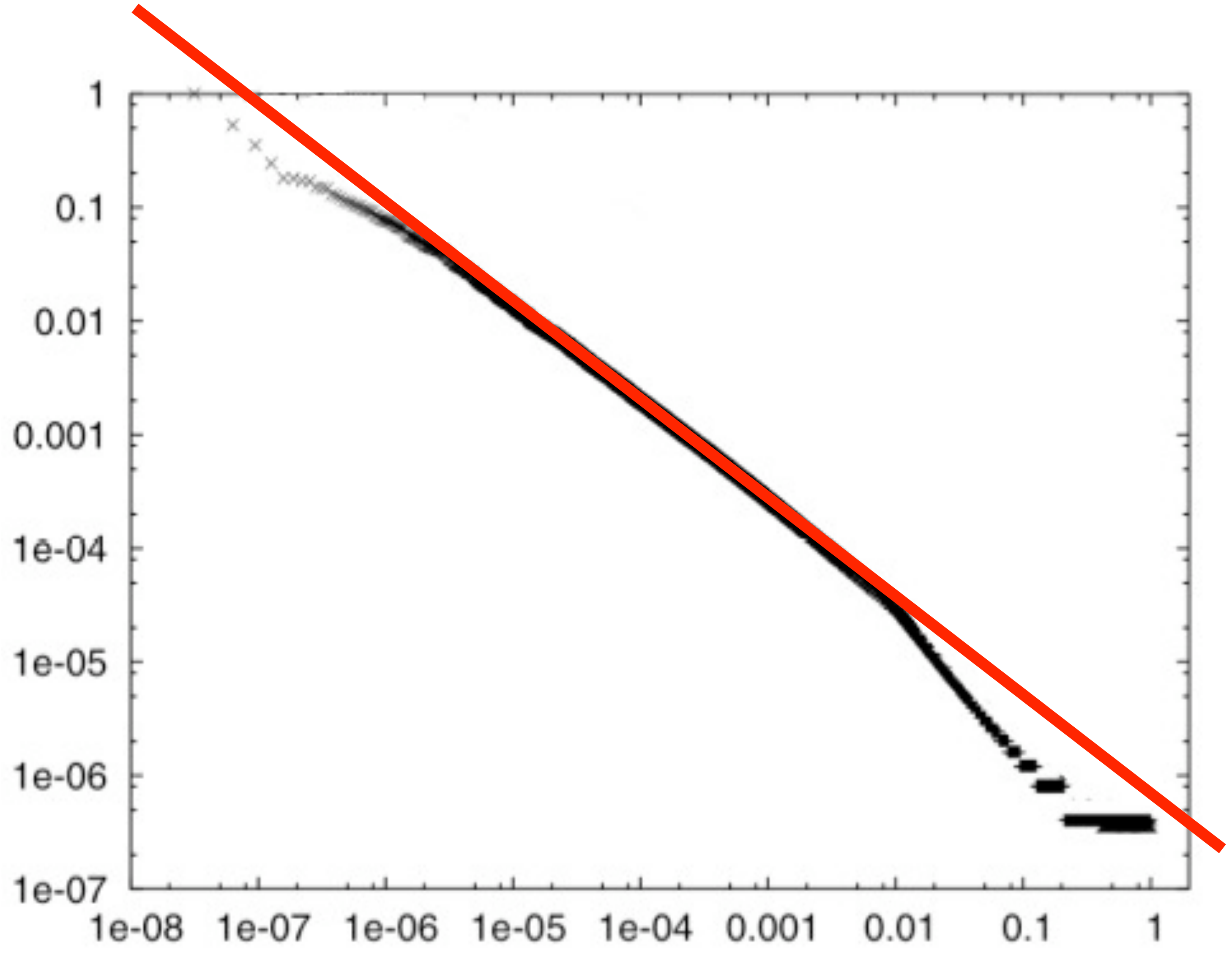
T. Fagni, R. Perego, F. Silvestri, and S. Orlando, "**Boosting the performance of web search engines: Caching and prefetching query results by exploiting historical usage data**," ACM Trans. Inf. Syst., vol. 24, no. 1, pp. 51–78, 2006.

Power-Law In Query Popularity: Excite



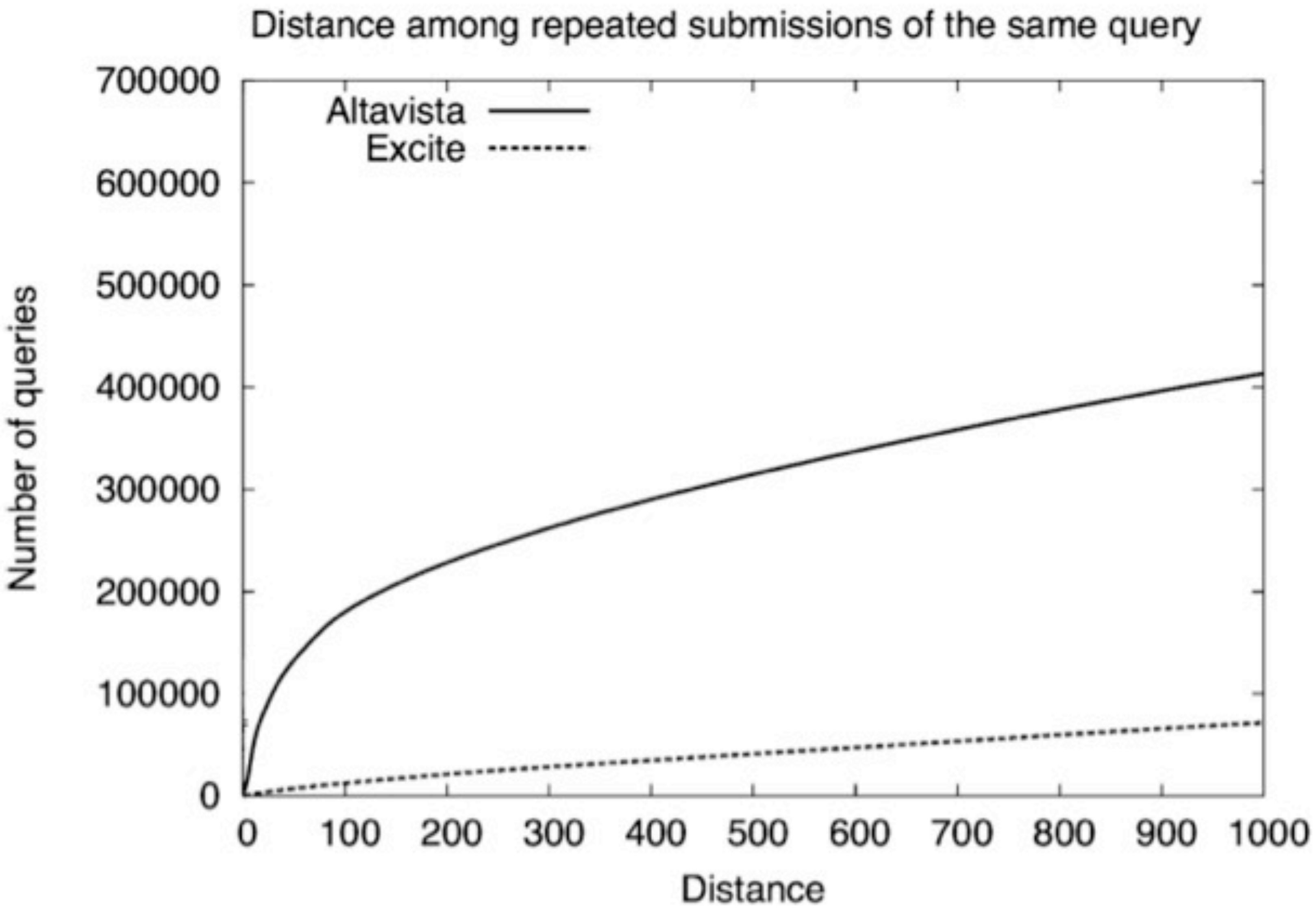
T. Fagni, R. Perego, F. Silvestri, and S. Orlando, "**Boosting the performance of web search engines: Caching and prefetching query results by exploiting historical usage data**," ACM Trans. Inf. Syst., vol. 24, no. 1, pp. 51–78, 2006.

Power-Law In Query Popularity: Yahoo!



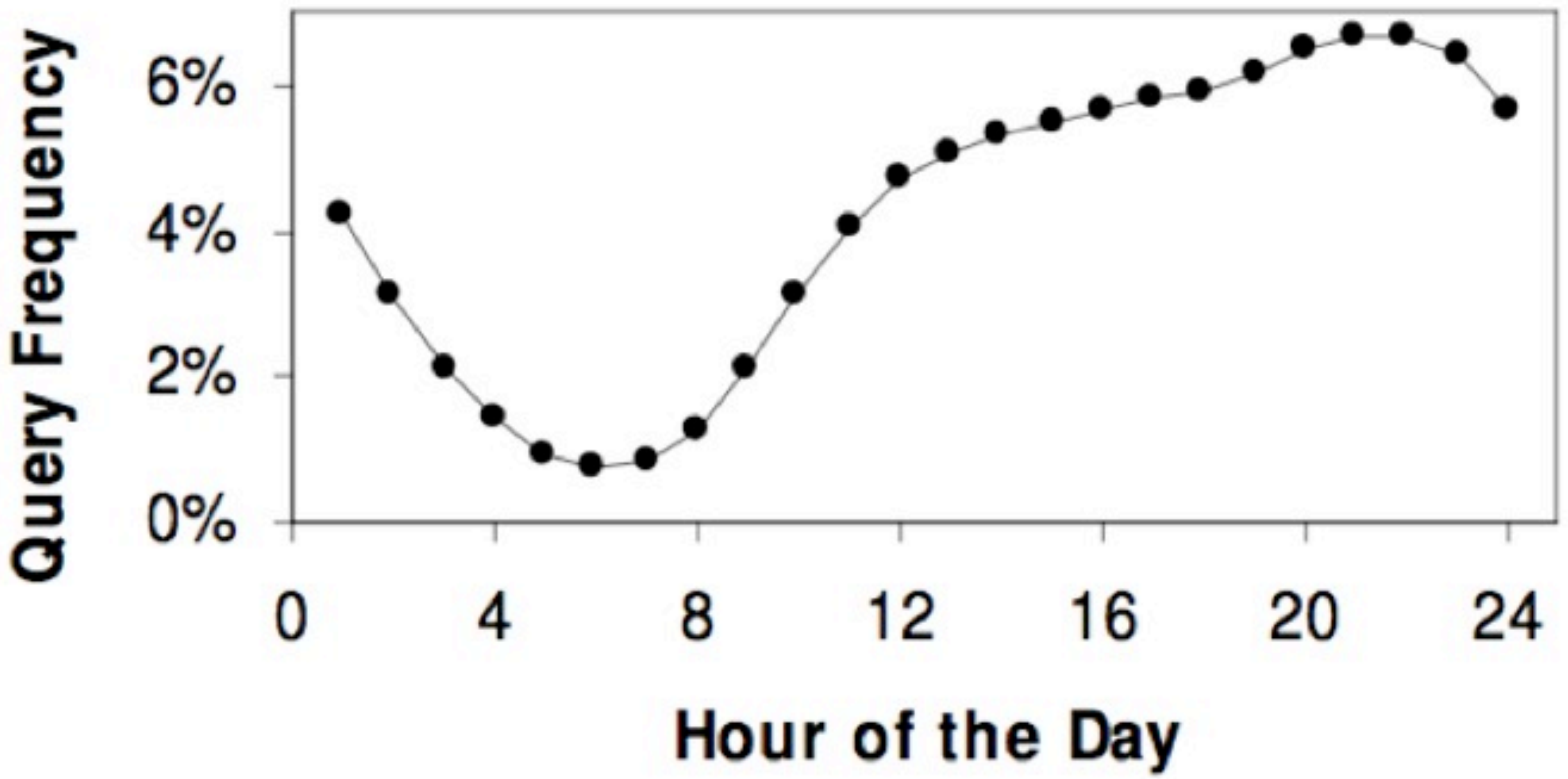
R. Baeza-Yates, A. Gionis, F. P. Junqueira, V. Murdock, V. Plachouras, and F. Silvestri, "**Design trade-offs for search engine caching**," ACM Trans. Web, vol. 2, no. 4, pp. 1–28, 2008.

Query Resubmission



T. Fagni, R. Perego, F. Silvestri, and S. Orlando, "**Boosting the performance of web search engines: Caching and prefetching query results by exploiting historical usage data**," ACM Trans. Inf. Syst., vol. 24, no. 1, pp. 51–78, 2006.

Frequency of Query Submission



S. M. Beitzel, E. C. Jensen, A. Chowdhury, O. Frieder, and D. Grossman, "T**emporal analysis of a very large topically categorized web query log**," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 2, pp. 166–178, 2007.

Query Statistics: Excite

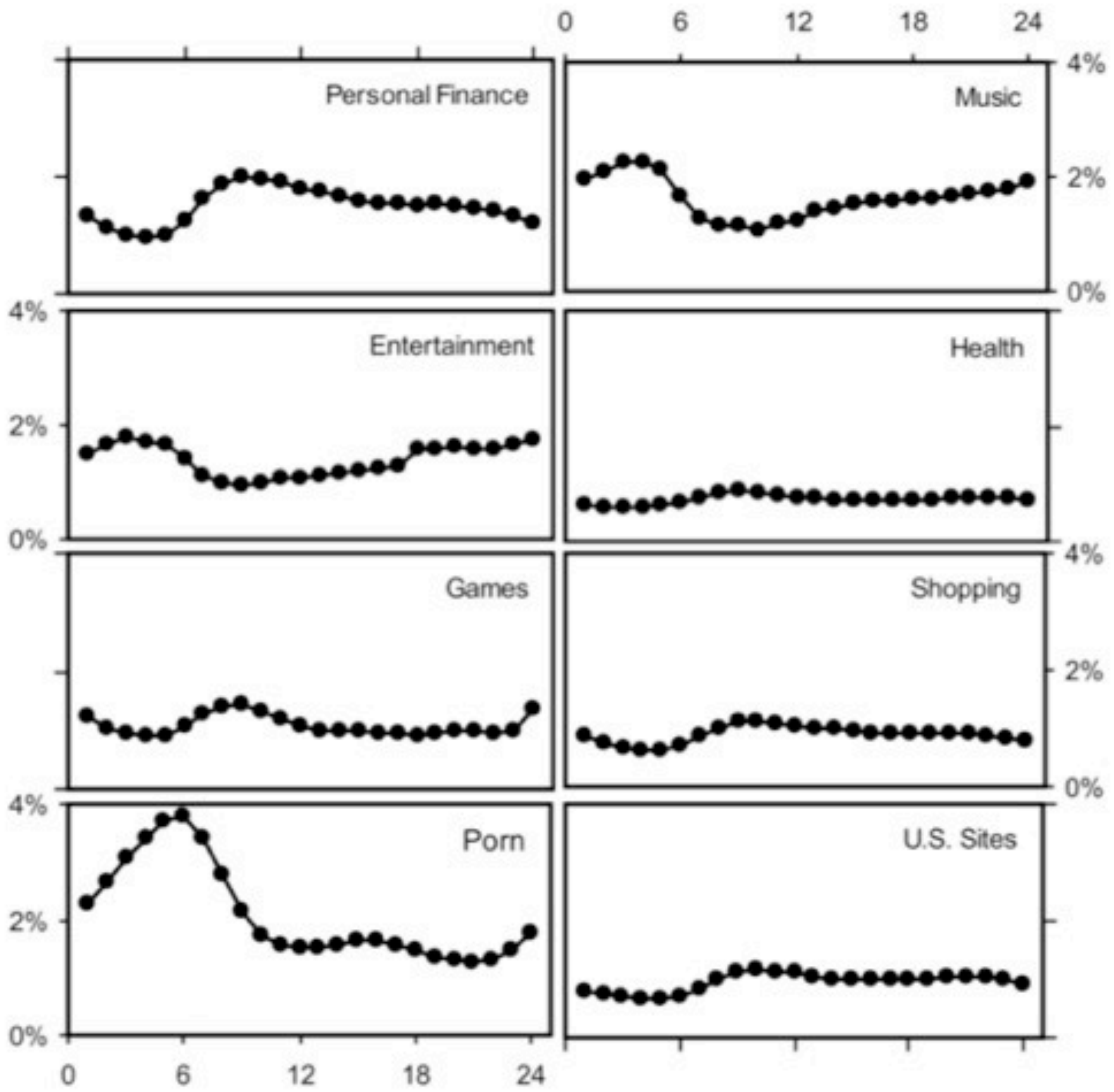
Characteristic	1997	1999	2001
Mean terms per query	2.4	2.4	2.6
Terms per query			
1 term			
2 terms			
3+ terms			
Mean queries per user	2.5	1.9	2.3

In 2008: 2.5 terms per query.

R. Baeza-Yates, A. Gionis, F. P. Junqueira, V. Murdock, V. Plachouras, and F. Silvestri, "**Design trade-offs for search engine caching**," ACM Trans. Web, vol. 2, no. 4, pp. 1–28, 2008.

A. Spink, B. J. Jansen, D. Wolfram, and T. Saracevic, "**From e-sex to e-commerce: Web search changes**," Computer, vol. 35, no. 3, pp. 107–109, 2002.

Hourly Topic Distribution

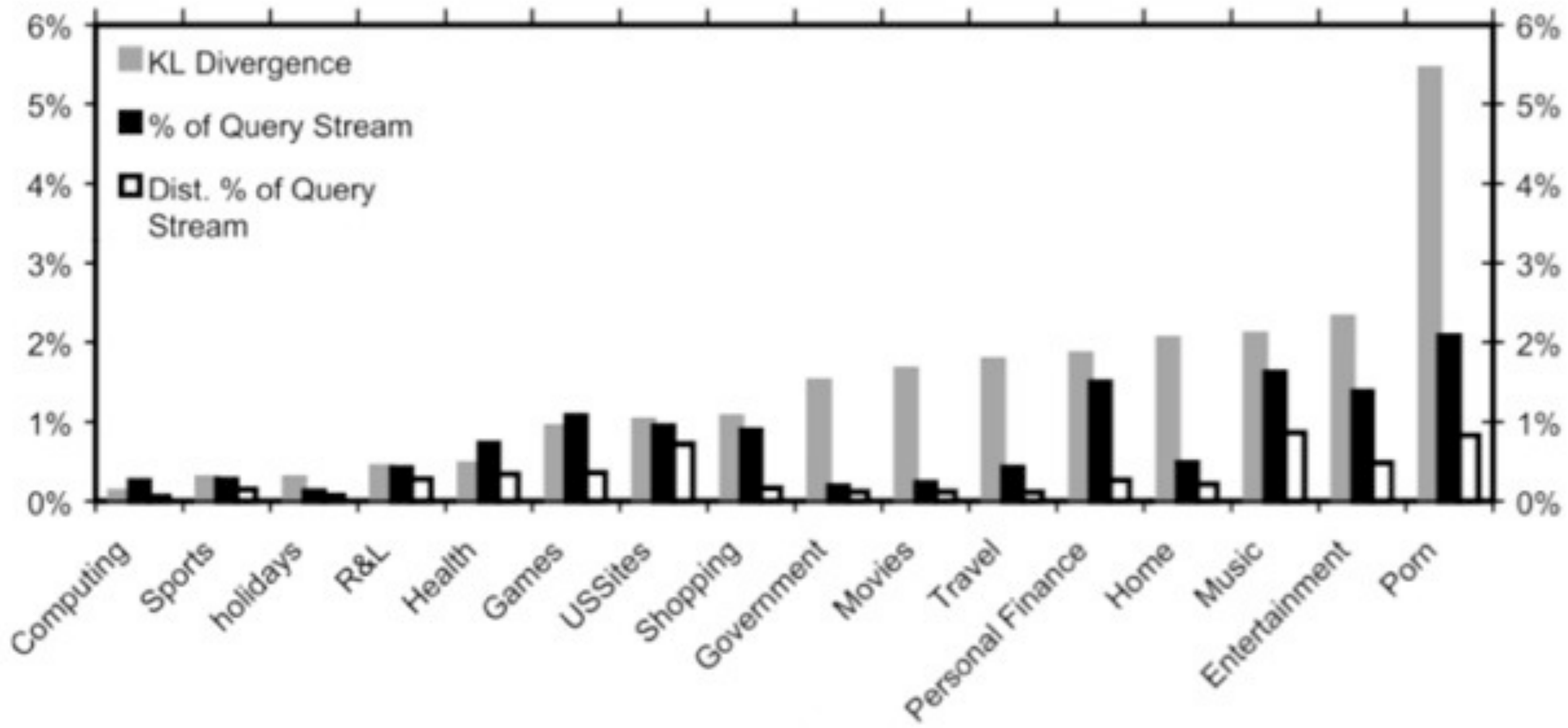


S. M. Beitzel, E. C. Jensen, A. Chowdhury, O. Frieder, and D. Grossman, "T**emporal analysis of a very large topically categorized web query log**," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 2, pp. 166–178, 2007.

Surprising Topics

$$D\left(p\left(q|t\right)\|p\left(q|c,t\right)\right)=\sum_{q}p\left(q|t\right)\log\frac{p\left(q|t\right)}{p\left(q|c,t\right)}$$

- KL-Divergence between the probab … and the actual topic observed.



S. M. Beitzel, E. C. Jensen, A. Chowdhury, O. Frieder, and D. Grossman, "**Temporal analysis of a very large topically categorized web query log**," J. Am. Soc. Inf. Sci. Technol., vol. 58, no. 2, pp. 166–178, 2007.

Summary of Query Statistics

- Web Search is different from traditional IR

	Traditional IR	Web Search
Query Length	6-9 (terms)	2-3 (terms)
Query Frequency	Zipf distribution	Zipf + skewed head and tail
# of SERPs viewed	about 10	1-2
Session Length	7-16 queries	1-2
Topics	Focused	(Highly) Diverse

Taxonomy of Web Search

- Navigational
  - Looking for a particular Web Site
- Informational
  - Willing to satisfy an information need
- Transactional
  - Willing to do some transactions through Web

A. Z. Broder, "**A taxonomy of web search**," SIGIR Forum, vol. 36, no. 2, pp. 3–10, 2002.

Navigational Queries

- American Airlines
- AA
- Google
- Yahoo
- CNN

They account for the 20 ~ 25% of the total queries.

Informational Queries

- High Dynamic Resolution Photos
- Escher
- Transfinite Numbers

They account for the 40 ~ 45% of the total queries.

Transactional Queries

- MP3
- Hotels Saint Petersburg
- Tickets for the Hermitage

They account for the 30 ~ 35% of the total queries.

Query Classification

- In the original Broder's paper they surveyed a group of volunteering Altavista users.
  - Some algorithmic classification has been done as well.
- More recent papers focused on automatic classification.

A Refined Taxonomy

Navigational | Informational | Transactional

Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy

Navigational

Informational

Resource

Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy

Navigational

Looking for a particular web site.

Informational

Willing to satisfy an information need

Resource

Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy

Navigational

Looking for a particular web site.

Informational

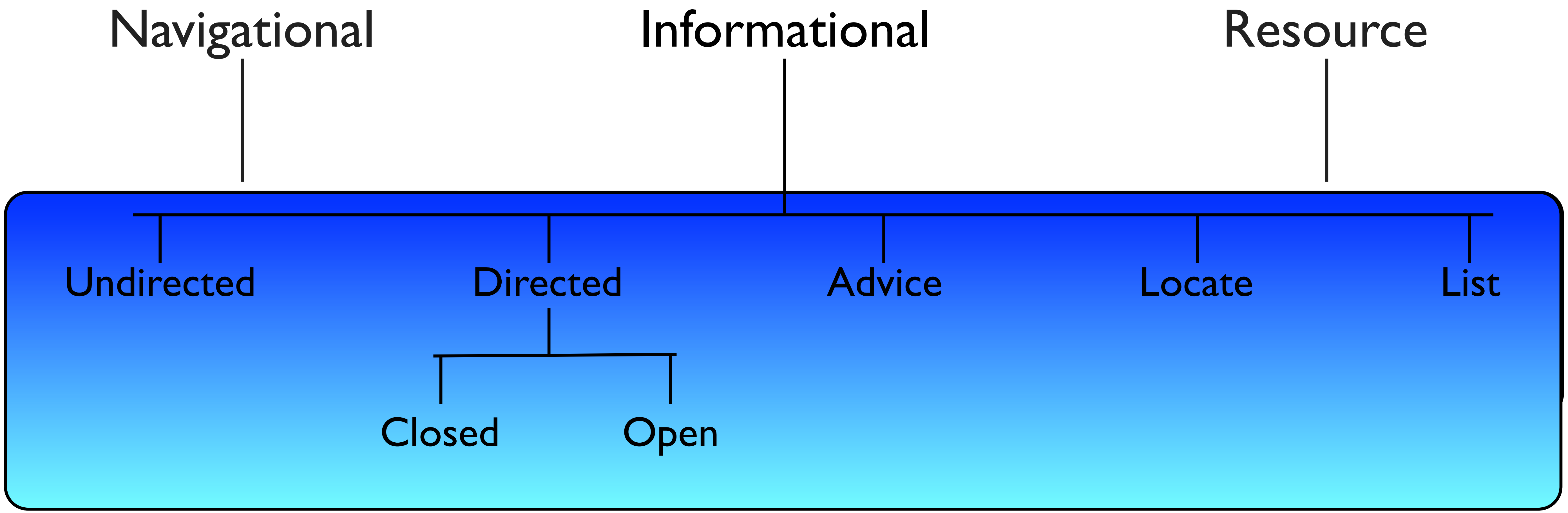
Willing to satisfy an information need

Resource

Looking for obtaining a resource (not information) available on the Web

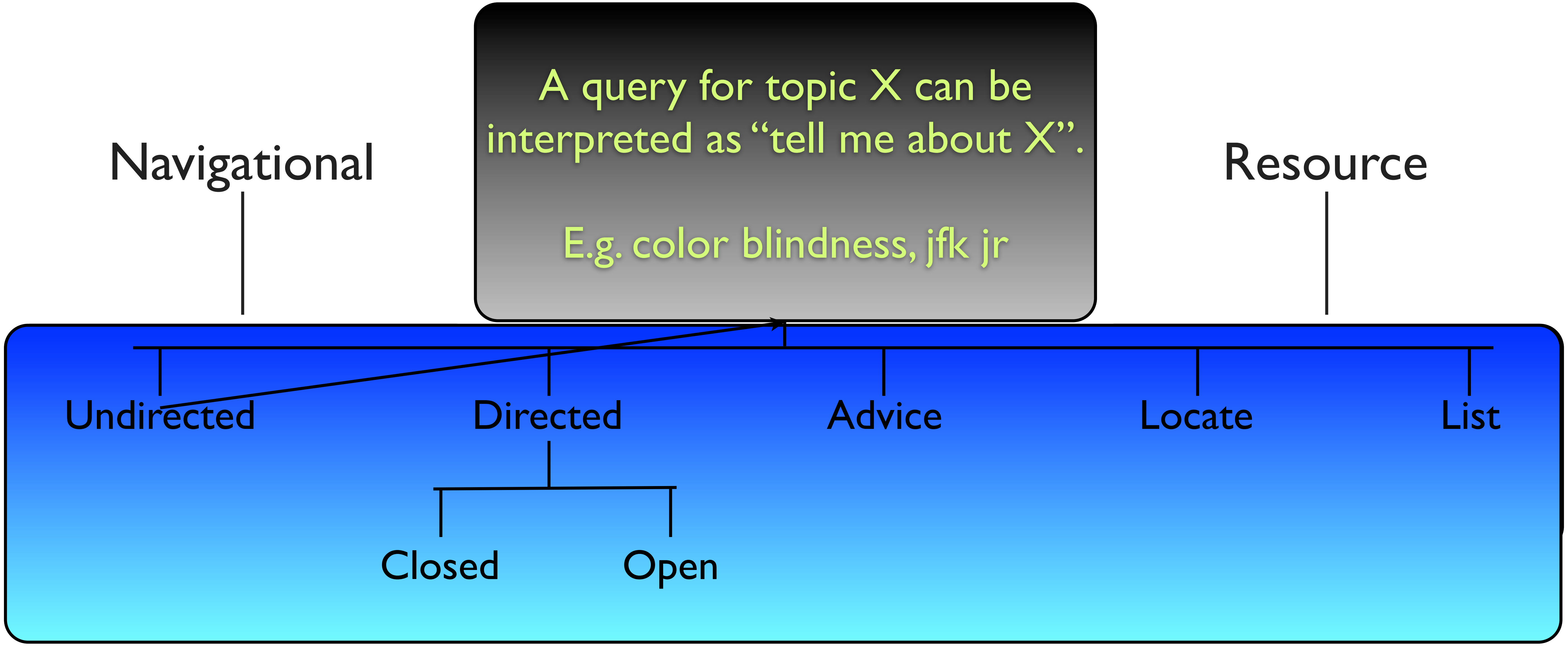
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



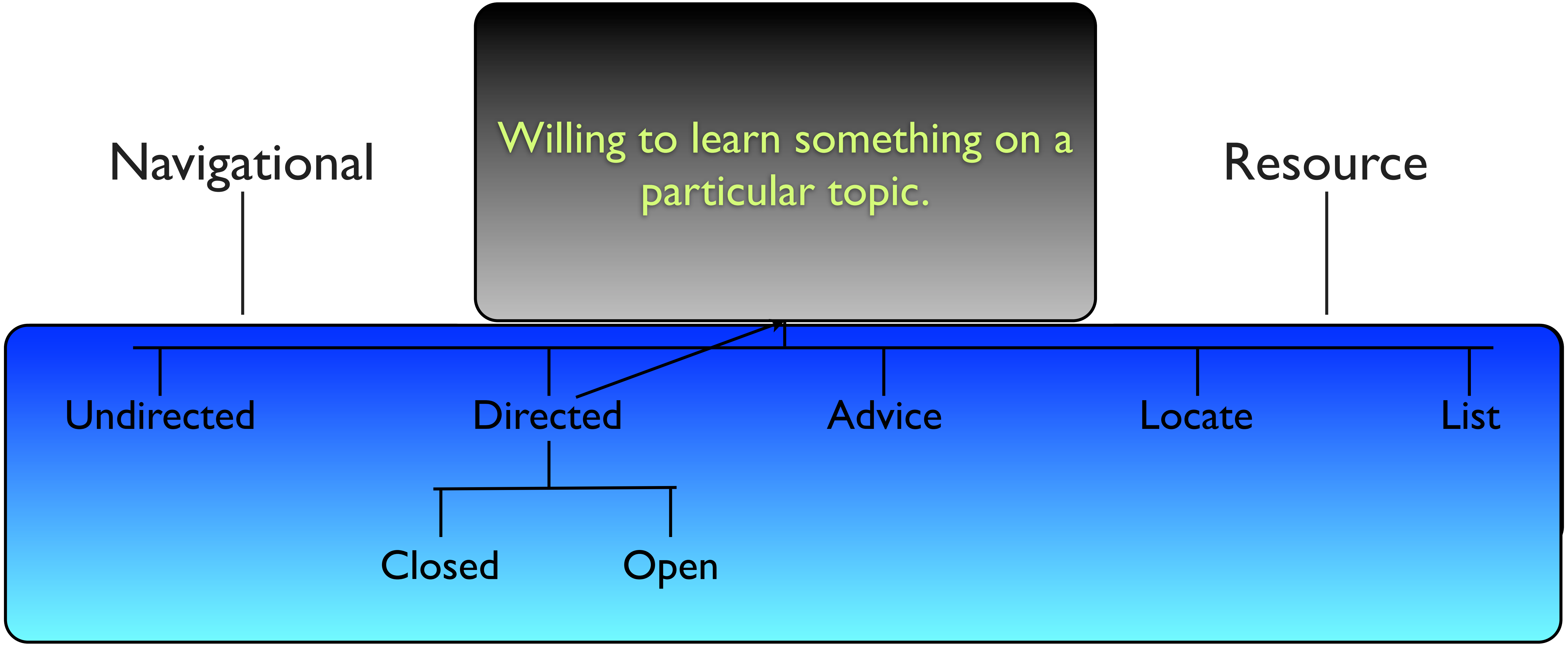
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



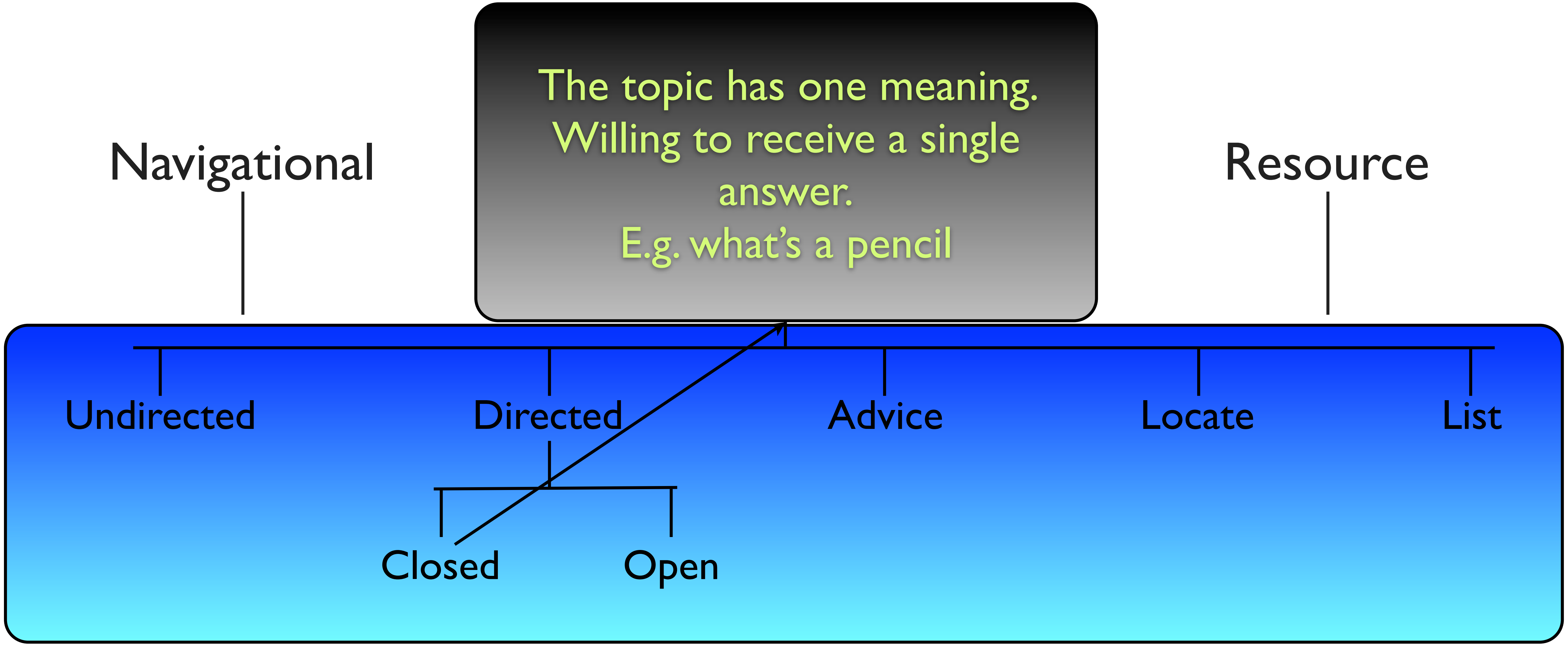
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



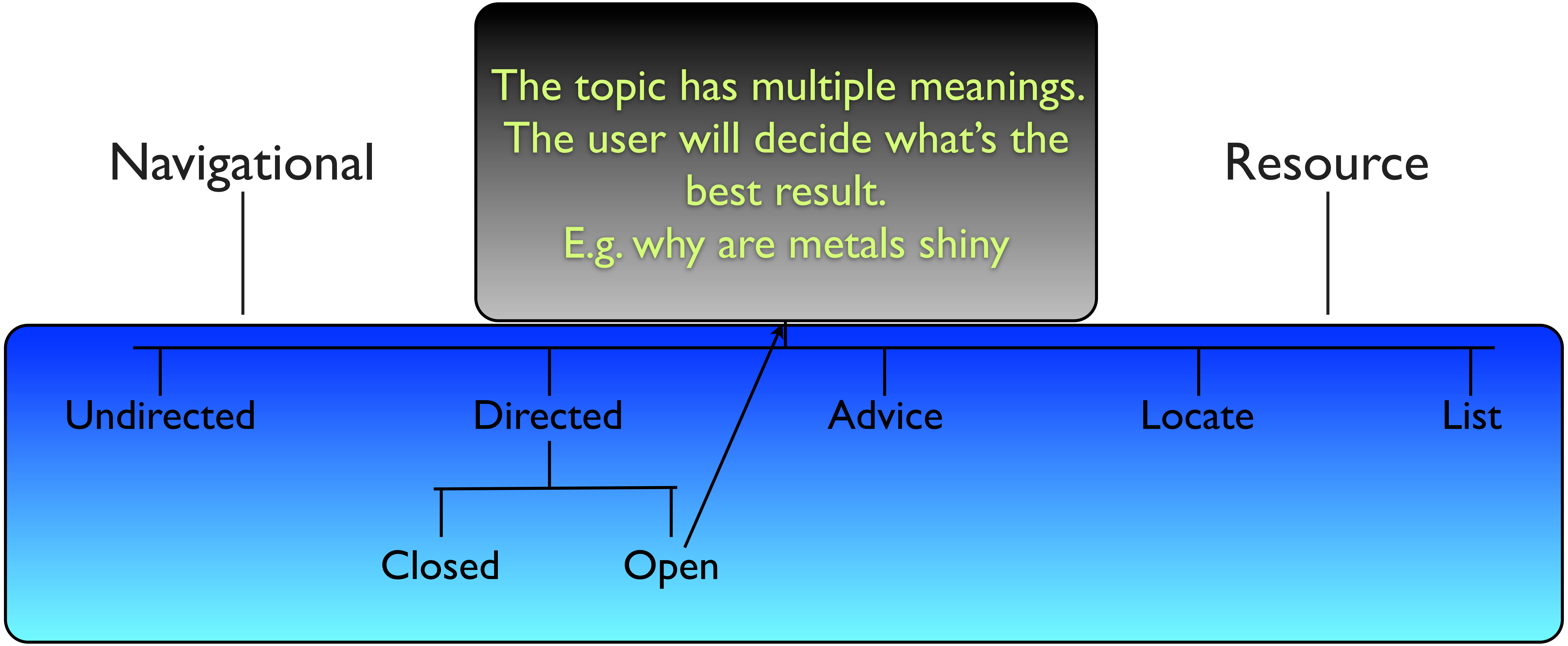
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



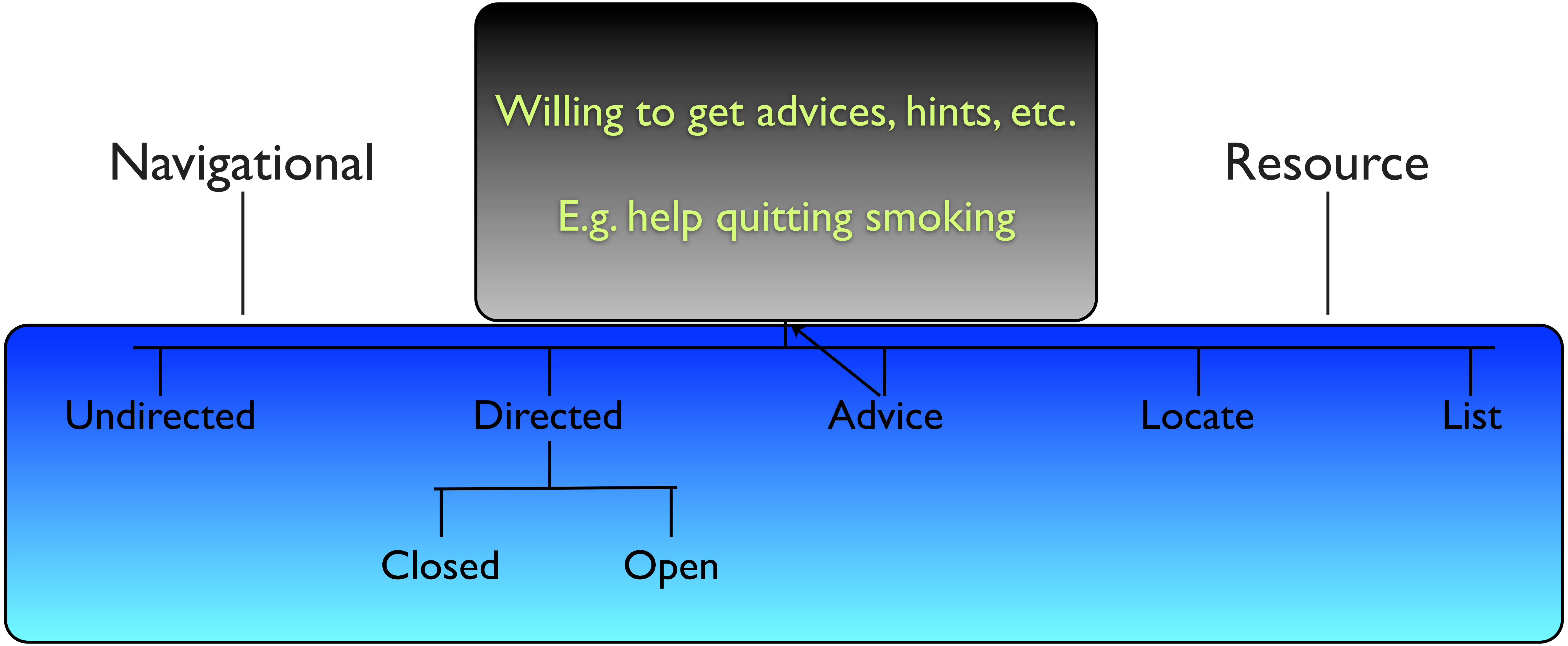
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



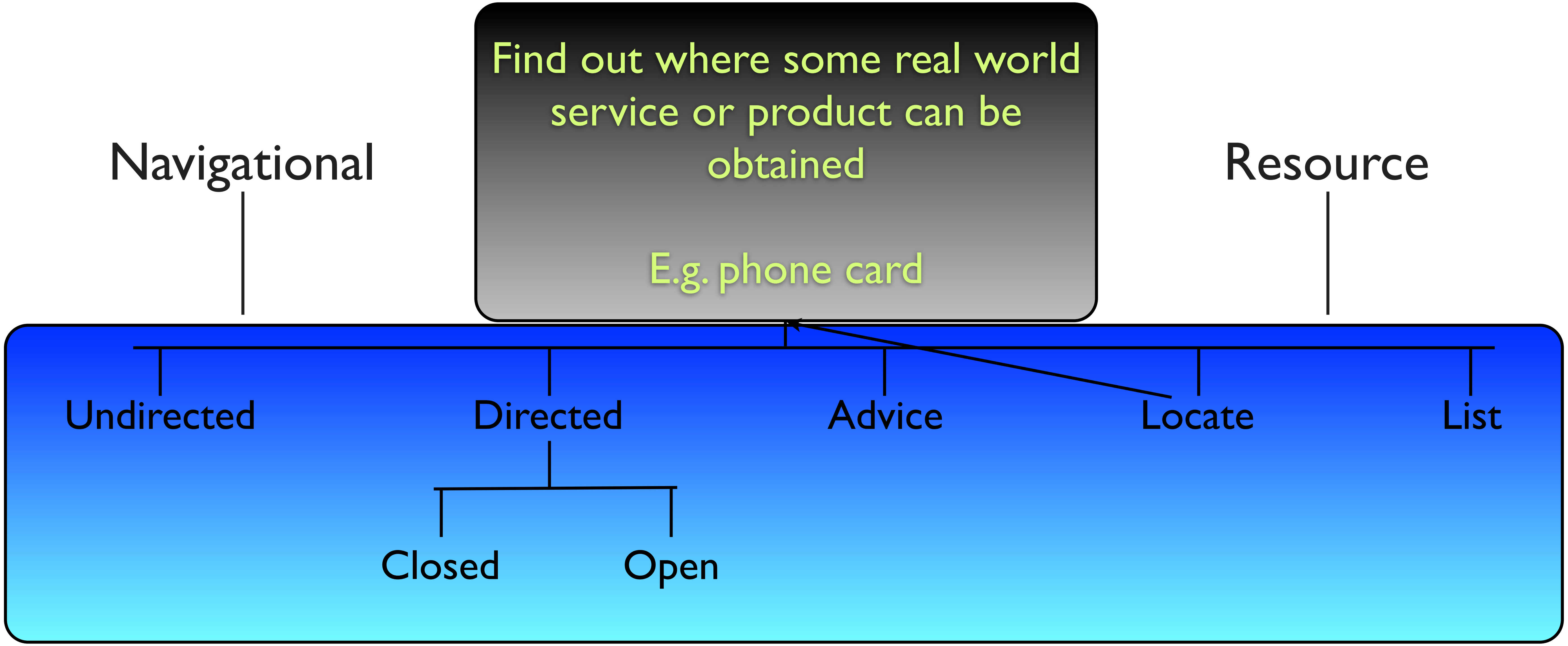
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy

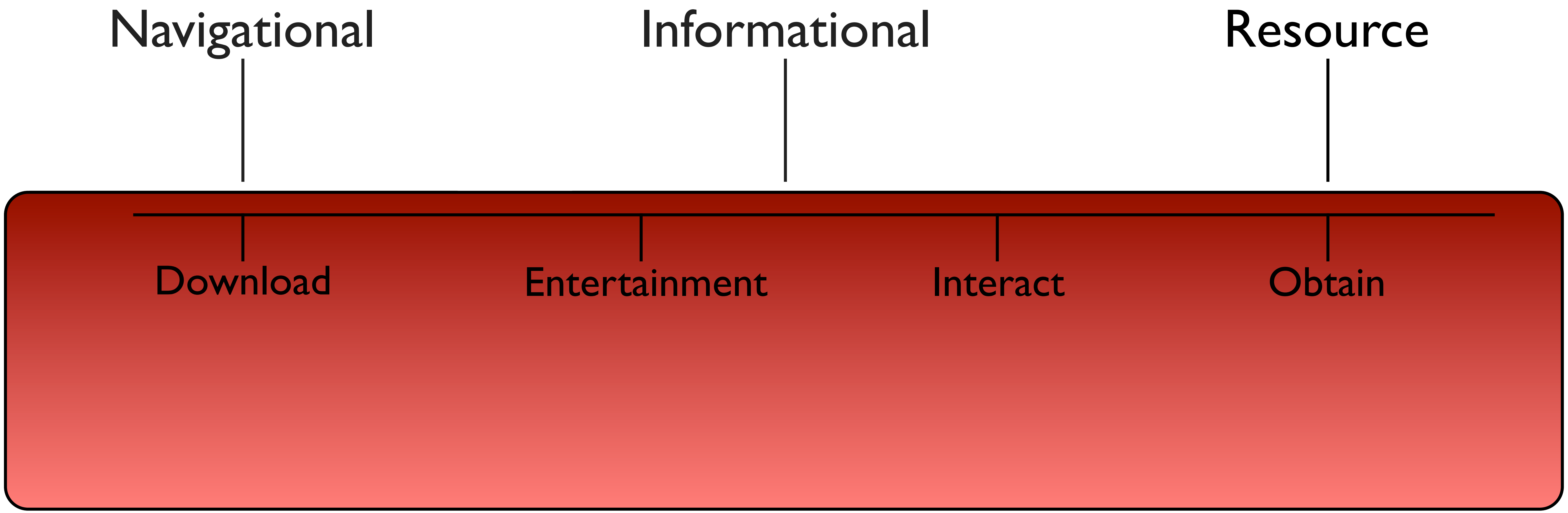


Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy

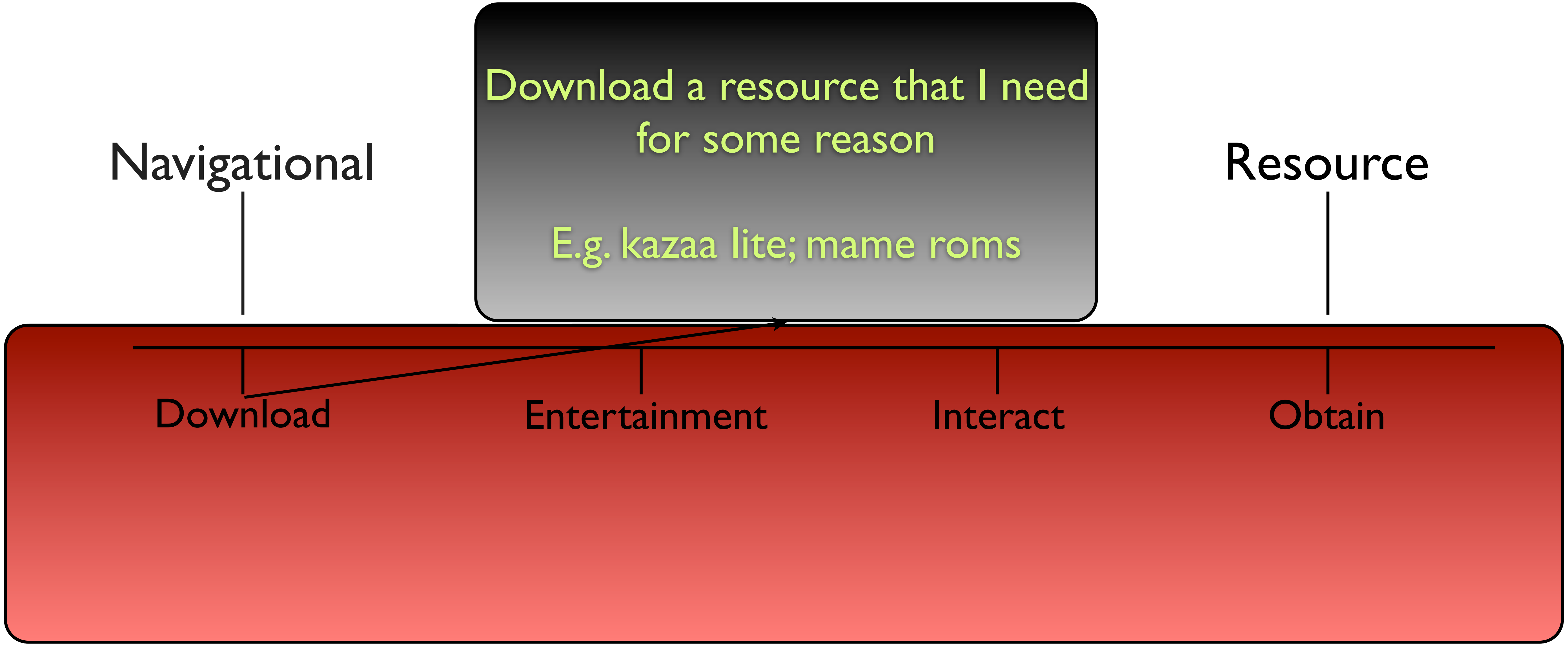
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



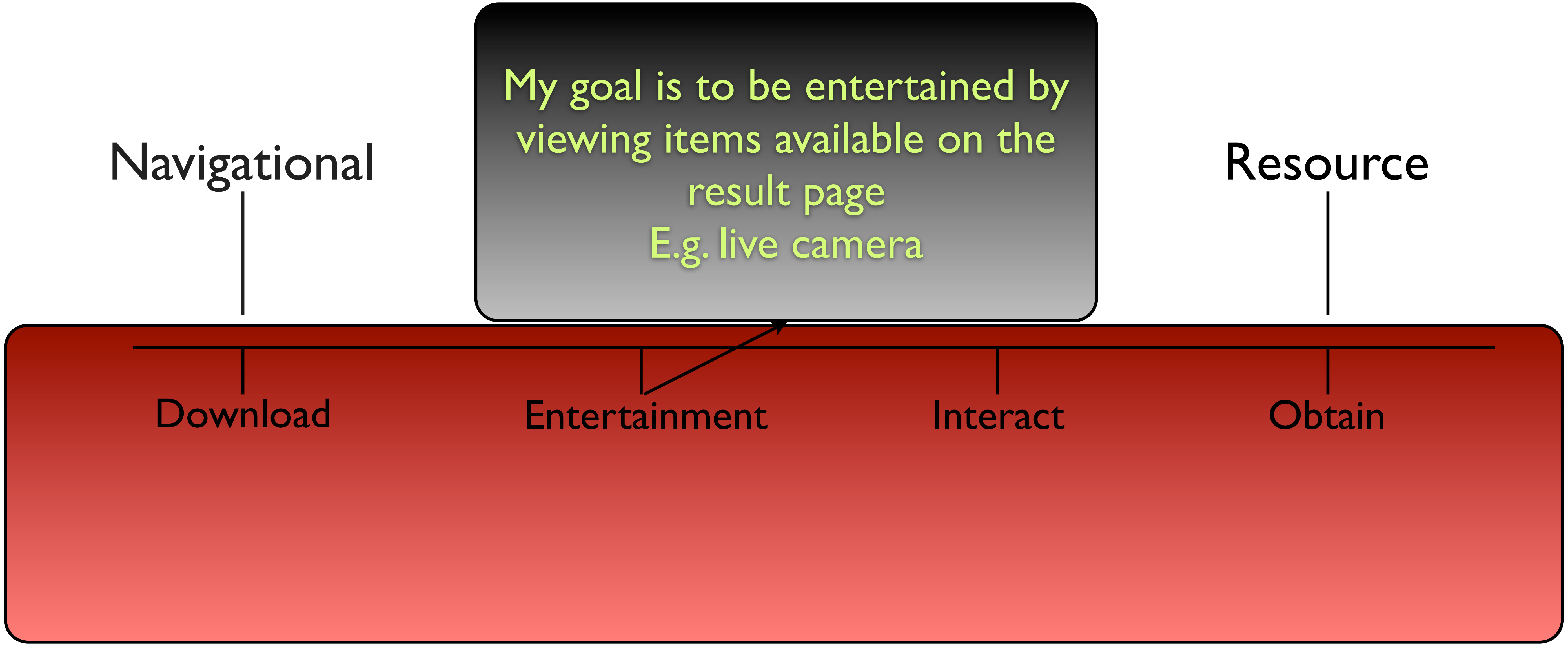
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



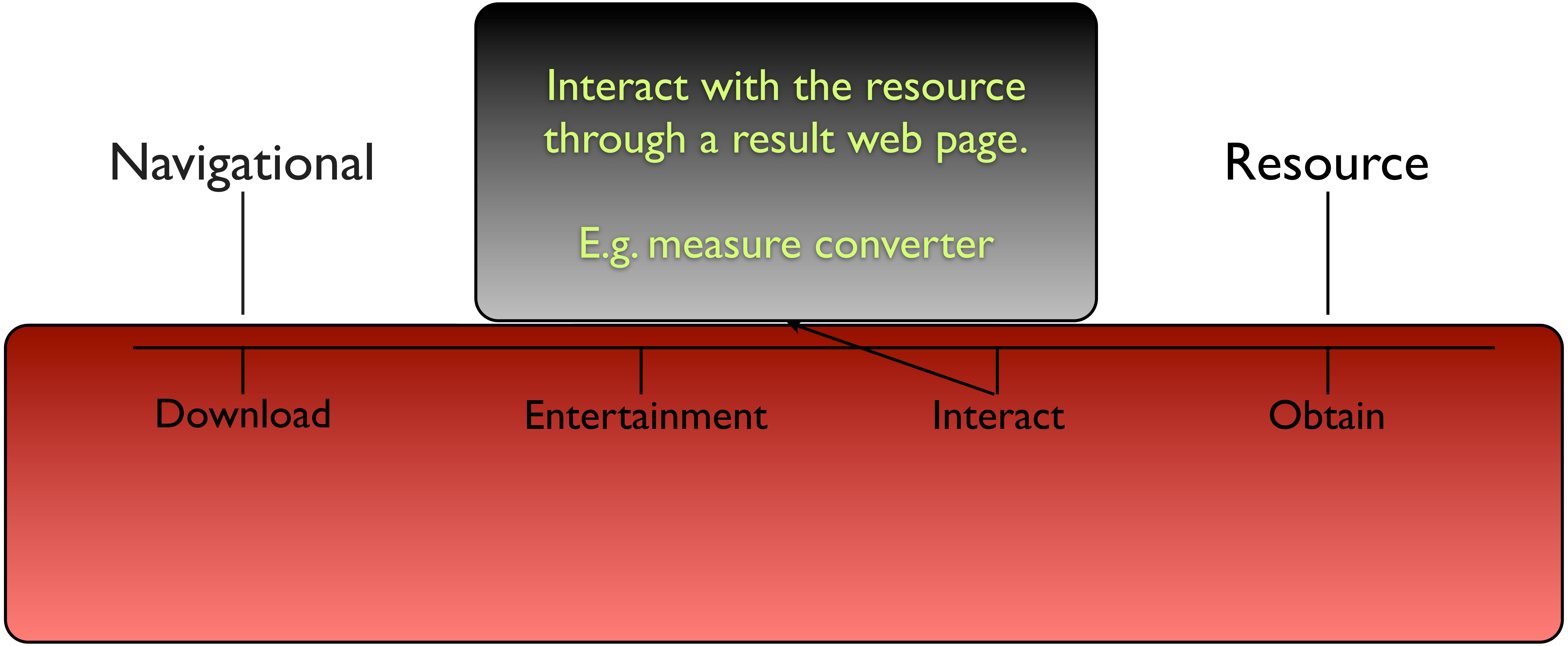
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



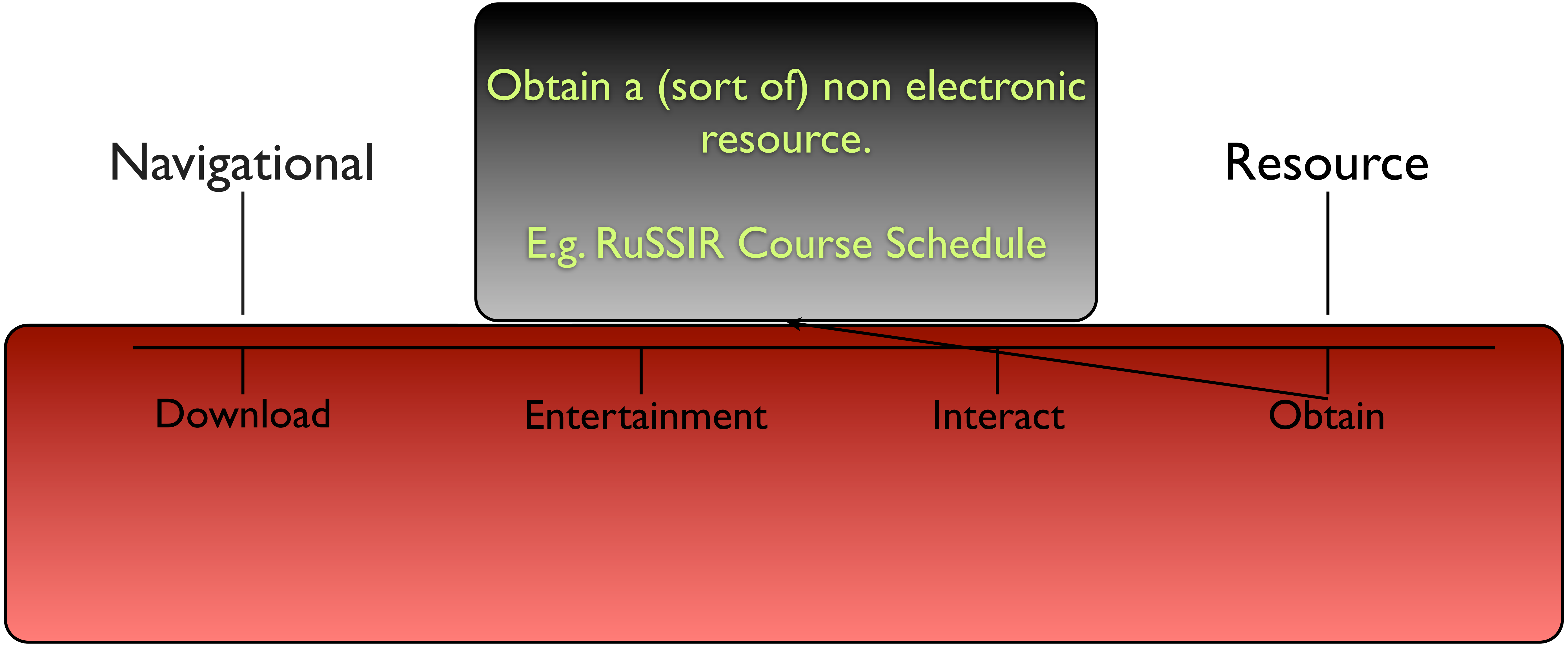
Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

A Refined Taxonomy



Rose, D. E. and Levinson, D. 2004. **Understanding user goals in web search**.
In Proceedings of WWW 2004 (New York, NY, USA, May 17 - 20, 2004). ACM, New York, NY, 13-19.

User Sessions

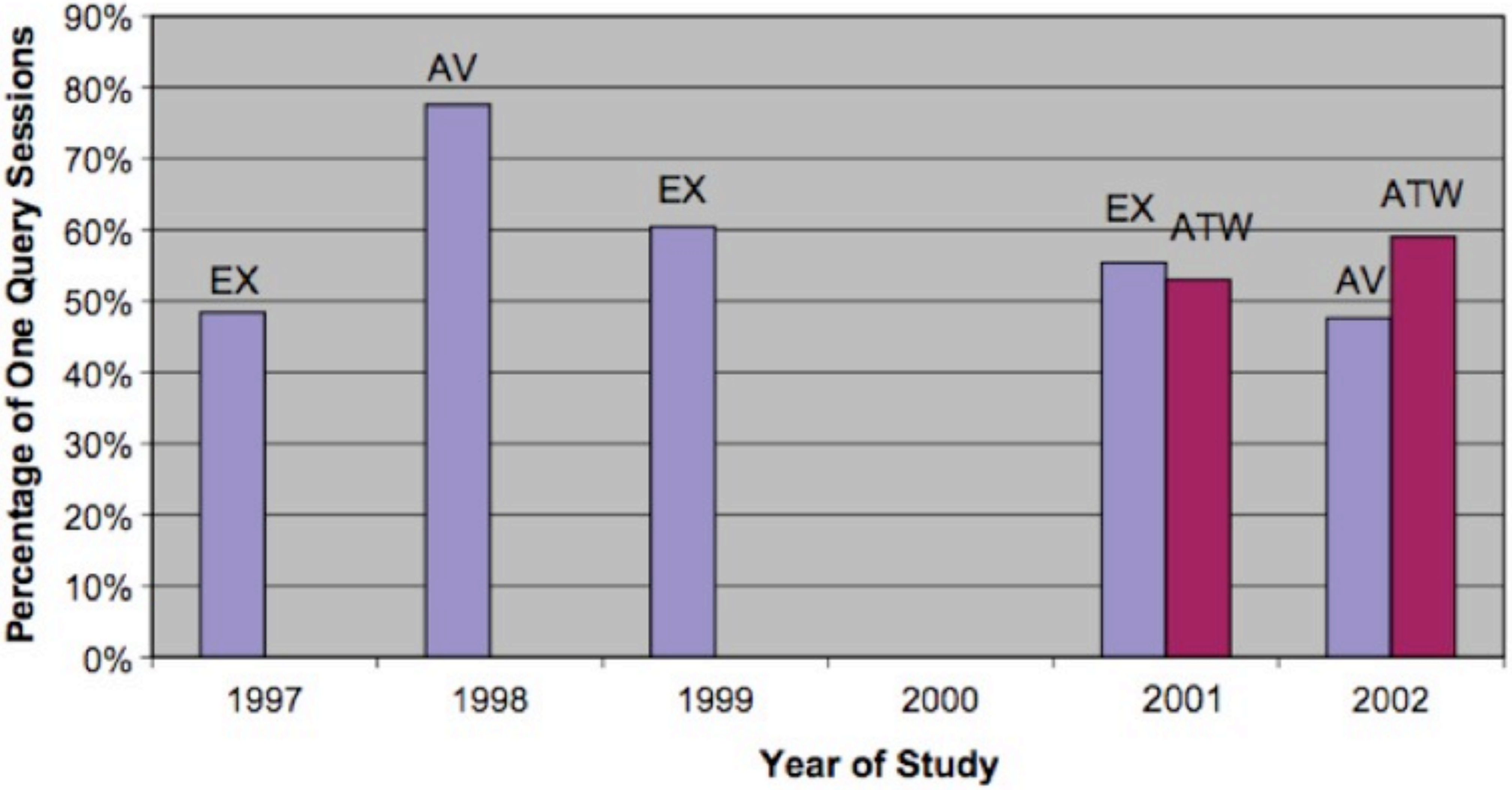
- A sequence of queries submitted by the same user is a user session.
- Usually a user is looking forward to satisfying a goal.

Typical Sessions

- Two queries of
  - two words, looking at
  - two answers page, doing
  - two clicks per page

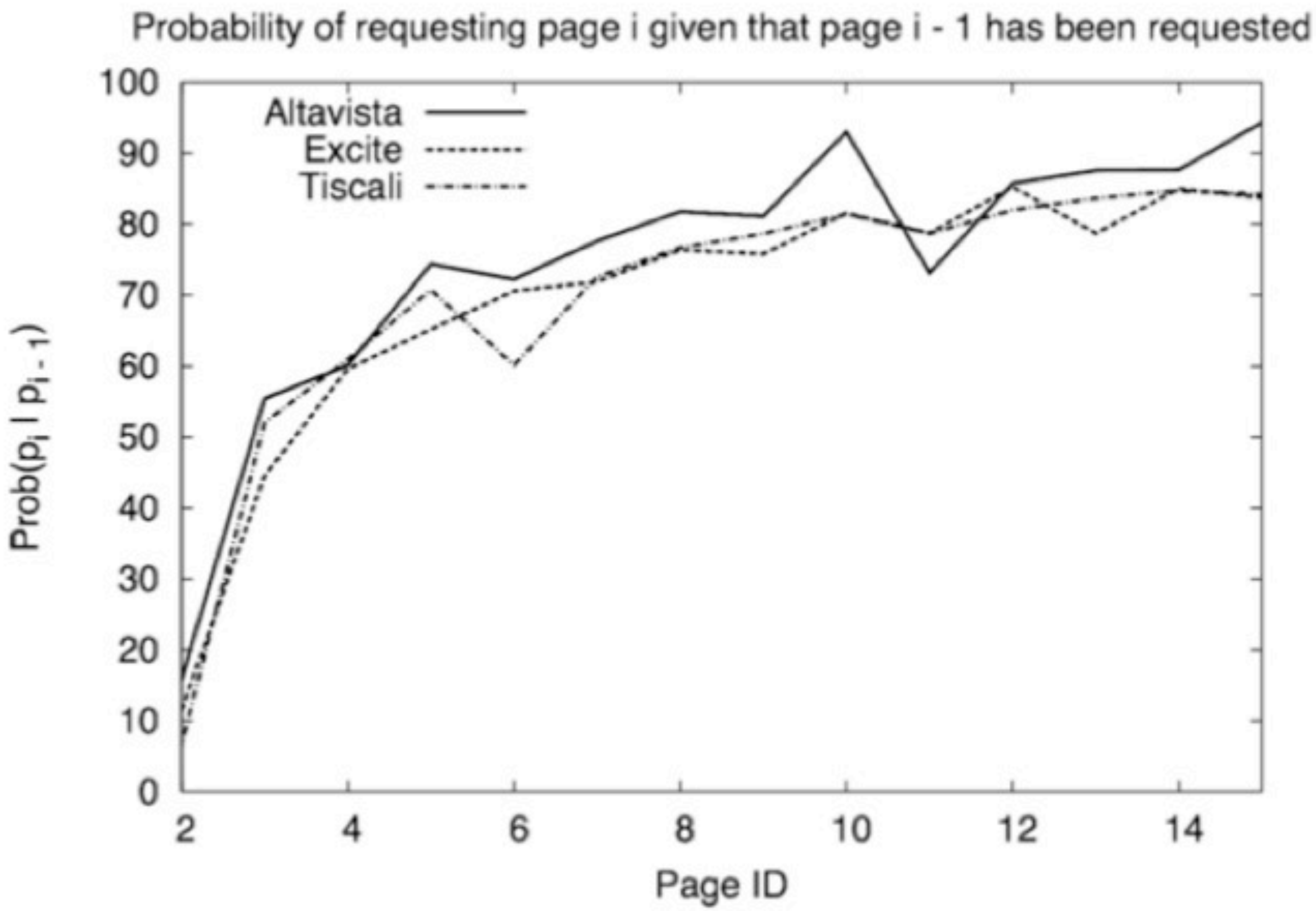
- Again: What is the goal?

Single Query Sessions



B. J. Jansen and A. Spink, "**How are we searching the world wide web? a comparison of nine search engine transaction logs**," Inf. Process. Manage., vol. 42, no. 1, pp. 248–263, 2006.

Multiple Query Sessions



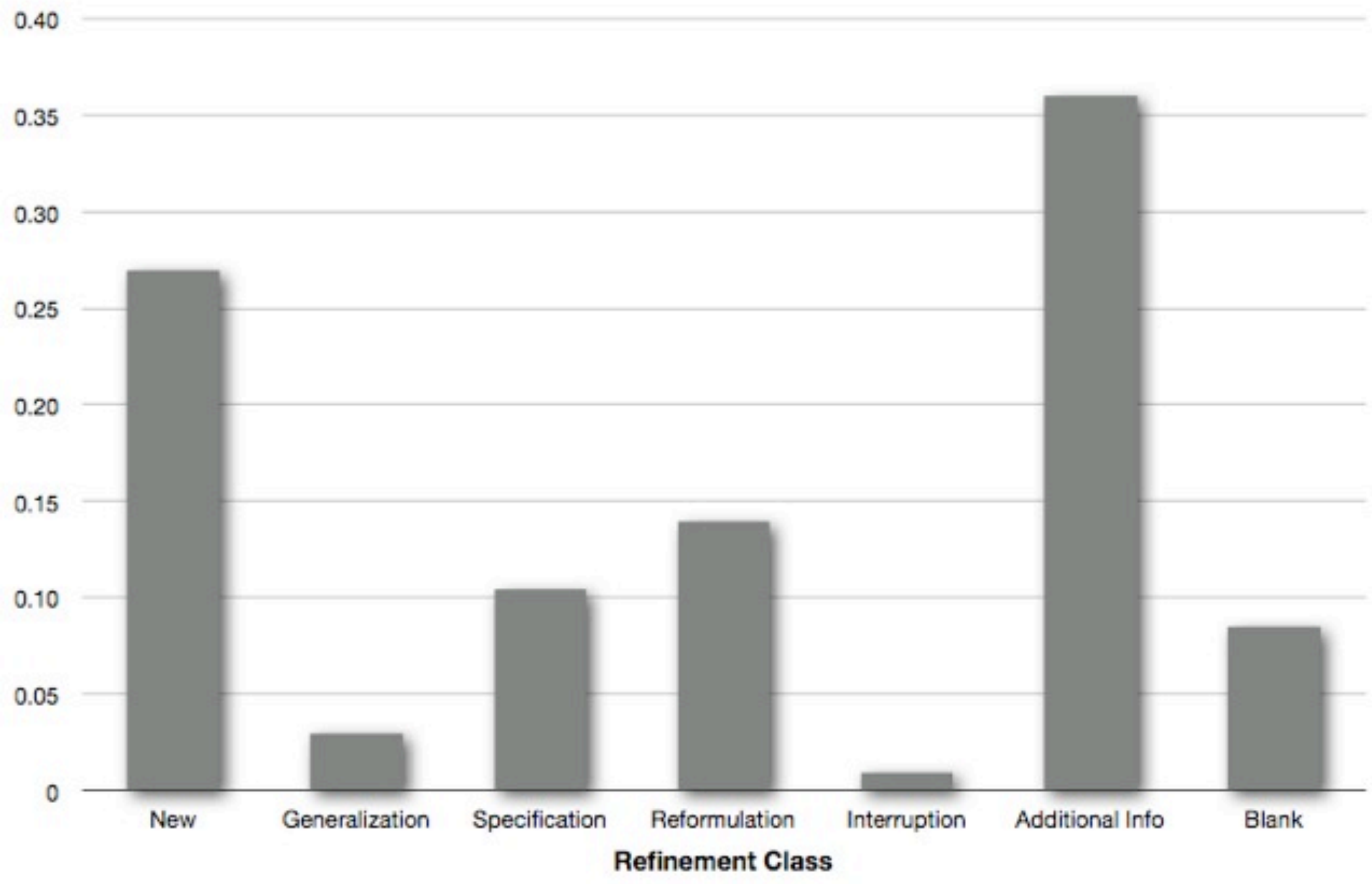
T. Fagni, R. Perego, F. Silvestri, and S. Orlando, "**Boosting the performance of web search engines: Caching and prefetching query results by exploiting historical usage data**," ACM Trans. Inf. Syst., vol. 24, no. 1, pp. 51–78, 2006.

Query Refinement

- New
- Generalization
- Specialization
- Reformulation
- Interruption
- Request for Additional Results
- Blank queries

T. Lau and E. Horvitz, "**Patterns of search: analyzing and modeling web query refinement**," in UM '99: Proceedings of the seventh international conference on User modeling, (Secaucus, NJ, USA), pp. 119–128, Springer-Verlag New York, Inc., 1999.

Query Refinement



T. Lau and E. Horvitz, "**Patterns of search: analyzing and modeling web query refinement**," in UM '99: Proceedings of the seventh international conference on User modeling, (Secaucus, NJ, USA), pp. 119–128, Springer-Verlag New York, Inc., 1999.

Query Resubmission

All queries: 13,060 queries (100%)	**Overlapping Click Queries** – 5072 queries (39%)			
	Equal Click Queries – 3777 (29%)			
	Single Identical Click 3737 (29%)	Multiple Identical Clicks 40 (< 1%)	Some Common Clicks 1295 (10%)	No Common Clicks 7988 (61%)
Equal Query Queries 4256 (33%)	**Navigational Queries** 3100 (24%)	36 (< 1%)	635 (5%)	485 (4%)
Different Query 8804 (67%)	637 (5%)	4 (< 1%)	660 (5%)	7503 (57%)

J. Teevan, E. Adar, R. Jones, and M. A. S. Potts, "**Information re-retrieval: repeat queries in yahoo's logs**," in SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, (New York, NY, USA), pp. 151–158, ACM, 2007.

Demographics of Web Search

- Ingmar Weber, Carlos Castillo: The demographics of web search. SIGIR 2010: 523-530
  - How does the web search behavior of ``rich'' and ``poor'' people differ?
  - Do men and women tend to click on different results for the same query?
  - What are some queries almost exclusively issued by African Americans?

Some Examples

Feature	Query	Value
Per-capita income k$	chris jordan	81k
	electric candle warmer	78k
	www.popsugar.com	75k
	ns4w.org	65k
below poverty line %	www.unitnet.com	26.4
	slaker	25.8
	kipasa	24.9
	www.tokbox.com	24.4
BA degree %	spencer stuart executive search	55.5
	insight venture partners	54.2
	federal circuit	53.2
	four seasons jackson hole	52.8

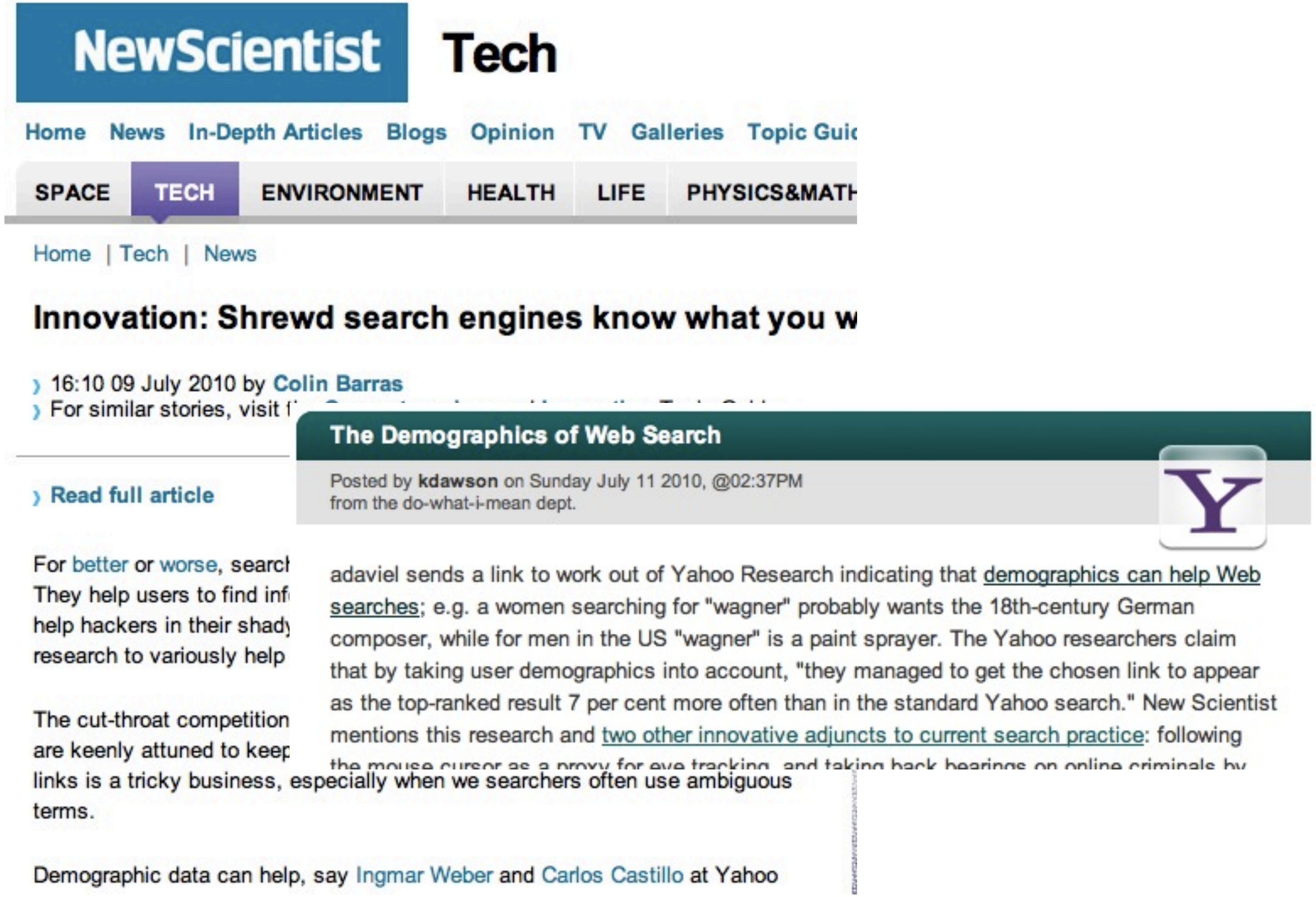
White %	pulloff.com	97.1
	central boiler wood furnace	96.2
	firewood processors	96.1
	midwest super cub	95.5
African Americ. %	trey songz bio	63.8
	def jam records address	58.4
	s2s magazine	58.1
	madinaonline	56.0
Asian %	sina	25.1
	big bang lyrics	24.3
	tvb series	24.2
	jay chou lyrics	23.5
Non-english lang. %	mis novelas favoritas	60.5
	sinonimos	59.2
	juegos para baby shower	54.5
	dichos mexicanos	54.3

Where Data Comes From?

- A subset of the query log data for US search traffic of the Yahoo! web search engine.
- Profile information (birth year, gender and ZIP code) provided by registered users.
- Publicly accessible demographic information for US ZIP codes, obtained in the 2000 census, and joined with the other data sources on the ZIP code (explicitly provided by users).

News Reported the Study

- Amongst others:
  - Slashdot
  - Newscientist



NewScientist Tech

Home News In-Depth Articles Blogs Opinion TV Galleries Topic Guic

SPACE TECH ENVIRONMENT HEALTH LIFE PHYSICS&MATH

Home | Tech | News

Innovation: Shrewd search engines know what you w

16:10 09 July 2010 by Colin Barras
For similar stories, visit t

Read full article

For better or worse, searcl
They help users to find inf
help hackers in their shad
research to variously help

The cut-throat competition
are keenly attuned to keep
links is a tricky business, especially when we searchers often use ambiguous
terms.

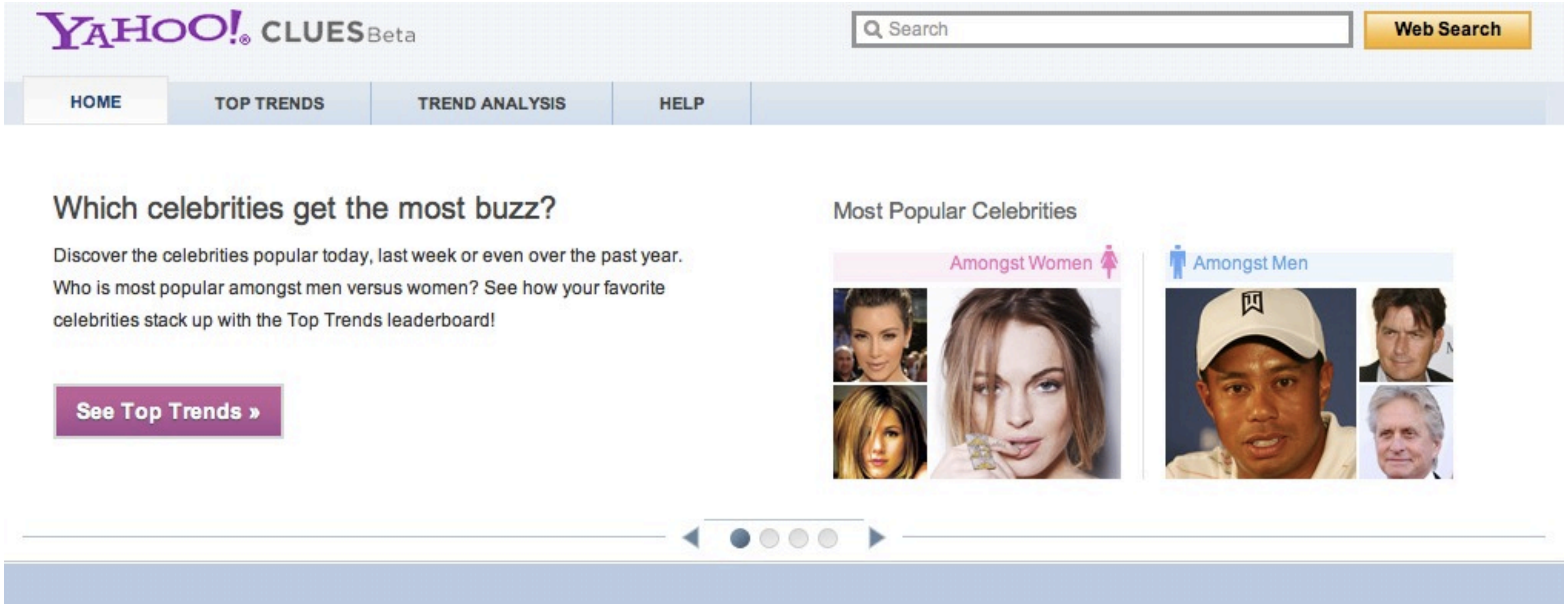
Demographic data can help, say Ingmar Weber and Carlos Castillo at Yahoo

The Demographics of Web Search

Posted by **kdawson** on Sunday July 11 2010, @02:37PM
from the do-what-i-mean dept.

adaviel sends a link to work out of Yahoo Research indicating that demographics can help Web searches; e.g. a women searching for "wagner" probably wants the 18th-century German composer, while for men in the US "wagner" is a paint sprayer. The Yahoo researchers claim that by taking user demographics into account, "they managed to get the chosen link to appear as the top-ranked result 7 per cent more often than in the standard Yahoo search." New Scientist mentions this research and two other innovative adjuncts to current search practice: following

Yahoo! Clues


YAHOO! CLUES Beta
Search
Web Search
HOME
TOP TRENDS
TREND ANALYSIS
HELP
Which celebrities get the most buzz?
Discover the celebrities popular today, last week or even over the past year. Who is most popular amongst men versus women? See how your favorite celebrities stack up with the Top Trends leaderboard!
See Top Trends »
Most Popular Celebrities
Amongst Women
Amongst Men

Questions?