Technical challenges in web advertising

Andrei Broder

Yahoo! Research



 This talk presents the opinions of the author. It does not necessarily reflect the views of Yahoo! Inc.

Advertising

What is it?



The two basic forms of advertising

- 1. Brand advertising
 - creates a distinct favorable image
- 2. Direct-marketing
 - Advertising that involves a "direct response": buy, subscribe, vote, donate, etc, **now or soon**





Sometimes both ...



©2006 Money Mailer LLC http://www.hotcoupons.com

H.O.T! Coupons Web Ad • 226-07-0082F

Web Advertising

There are lots of ads on the web ...







Textual: Sponsored Link:

Text-based ads usually driven by a keyword search

Sign in



Web	<u>lmages</u>	Groups	News	Froogle	<u>Maps</u>	mor	e »	
sigir							Search	Advanced Search Preferences

Web

Results 1 - 10 of about 1,540,000 for sigir (0.07 seconds)

SIGIR: Information Retrieval

"Addresses issues ranging from theory to user demands in the application of computers to the acquisition,...

www.acm.org/sigir/ - 7k - Cached - Similar pages

SIGIR • 2005

SIGIR is the major international forum for the presentation of new research ... The 28th Annual International ACM **SIGIR** Conference will be held at the ...

www.dcc.ufmg.br/eventos/sigir2005/ - 14k - Cached - Similar pages

Special Inspector General for Iraq Reconstruction: SIGIR Homepage

SIGIR, the successor to the Coalition Provisional Authority Inspector General (CPA-IG), ... **SIGIR** oversight is accomplished via independent audits, ...

www.sigir.mil/ - 16k - Cached - Similar pages

Sponsored Links

Sigir

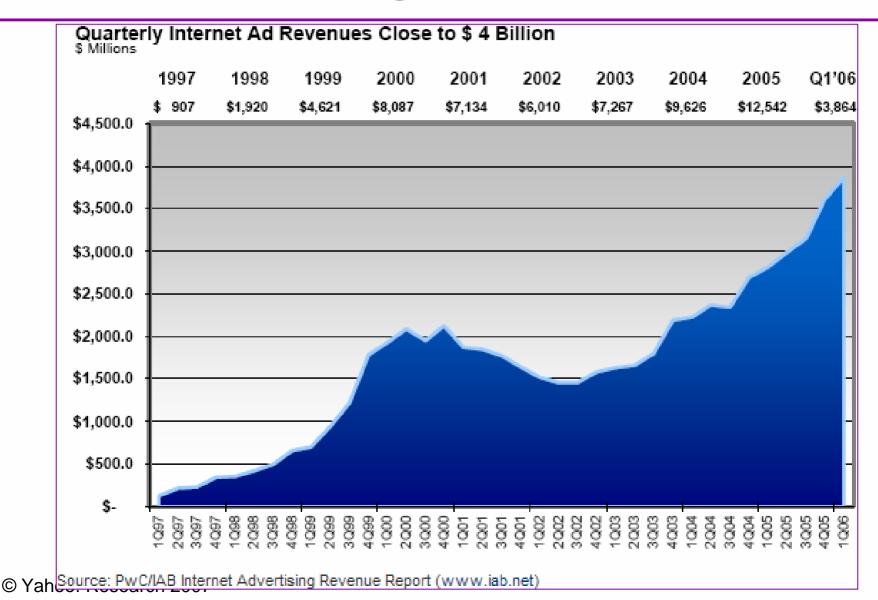
Whatever you're looking for you can get it on eBay. www.eBay.com

Sigir 2006 Tutorial

Introduction to Web Advertising Seattle, August 6, 2006 www.sigir2006.org/tutorials.asp



Ad revenue stats from Internet Advertising Board





- 1. Ads driven by search keywords "sponsored search" (a.k.a. "keyword driven ads", "paid search", etc)
- 2. Ads directly driven by the content of a web page "context match" (a.k.a."context driven ads", "contextual ads", etc)

Both types are heavily related to web search



A sponsored search ad

Sian in



Web	<u>lmages</u>	Groups	News	Froogle	<u>Maps</u>	mor	e »	
sigir							Search	Advanced Search Preferences

Web

Results 1 - 10 of about 1.540,000 for sigir (0.07 seconds)

SIGIR: Information Retrieval

"Addresses issues ranging from theory to user demands in the application of computers to the acquisition....

www.acm.org/sigir/ - 7k - Cached - Similar pages

SIGIR • 2005

SIGIR is the major international forum for the presentation of new research ... The 28th Annual International ACM SIGIR Conference will be held at the ...

www.dcc.ufmg.br/eventos/sigir2005/ - 14k - Cached - Similar pages

Special Inspector General for Iraq Reconstruction: SIGIR Homepage

SIGIR, the successor to the Coalition Provisional Authority Inspector General (CPA-IG). ... SIGIR oversight is accomplished via independent audits, ...

www.sigir.mil/ - 16k - Cached - Similar pages

Sponsored Links

Sigir

Whatever you're looking for you can get it on eBay. www.eBav.com

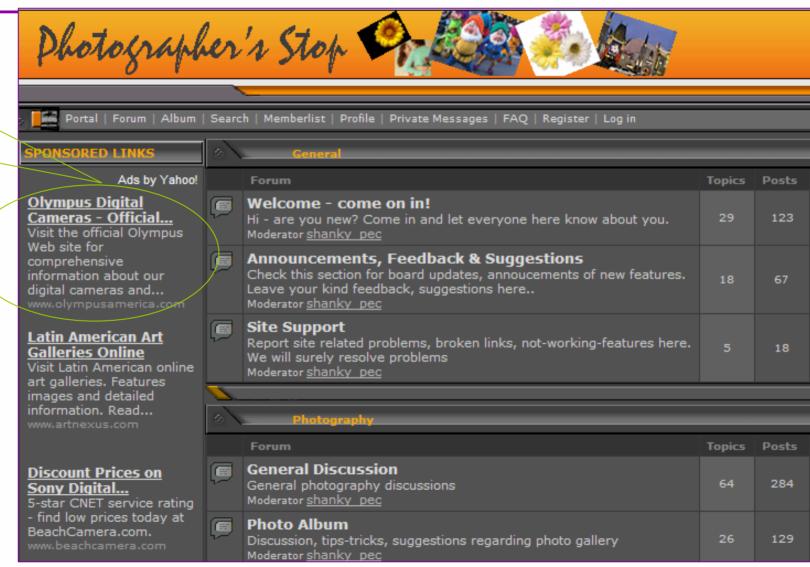
Sigir 2006 Tutorial

Introduction to Web Advertising Seattle, August 6, 2006 www.sigir2006.org/tutorials.asp



A content match ad

Content match ad





Web & Search Advertising business

- All large players (Google, Yahoo, MSN) now own:
 - 1. Web search engine
 - 2. Advertising platform + network
 - They are both a search engine and an ad agency
- Google
 - Own engine for keyword ads (Adwords)
 - Bought Applied Semantics (context ads, Adsense)
- Yahoo
 - Bought Inktomi, AltaVista, AlltheWeb, Overture
 - Search ← Inktomi derived
 - Ads ← Overture (keyword, context)
- MSN
 - Used search from Inktomi + own corpus. Has now its own
 - Used ads from Overture, has now own platform ("AdCenter" for keywords from May 2006, announced "ContentAds" beta in summer)

14



Concepts

- –CPM = cost per thousand impressions
 - Typically used for graphical/banner ads
- -CPC = cost per click
- –CPT/CPA = cost per transaction/action a.k.a. referral fees or affiliate fees
 - Typically used for shopping ("buy from our sponsors"), travel, etc



We advertised a tutorial:



Sponsored Links

WWW2006 Tutorial

Foundations and challenges of web advertising -- Edinburgh, May 24 www2006.org/tutorials/#T01

- Started late ...
- Spent < \$10.00 total

© Yahoo! Research 2007

16



Our new ad in Google

Sian in



Groups News Frooale Maps more » Advanced Search Search

Preferences

Web

Results 1 - 10 of about 197,000 for andrei broder. (0.05 seconds)

House ad

DBLP: Andrei Z. Broder

62. Andrei Z. Broder: Introduction: The Fourth International Workshop on ... 3, Andrei Z. Broder: A Provably Secure Polynomial Approximation Scheme for the ...

www.informatik.uni-trier.de/~lev/ db/indices/a-tree/b/Broder:Andrei Z=.html - 59k Cached - Similar pages

Andrei Broder Joins Yahoo

Andrei Broder, former vice president of research at AltaVista and until recently Distinguished Engineer & CTO, IBM Research, is joining Yahoo as research ... blog.searchenginewatch.com/blog/051118-122544 - 65k - Cached - Similar pages Sponsored Links

Work at Google

Google is hiring expert compute scientists and software developers! www.google.com/jobs

Sigir 2006 Tutorial

Introduction to Web Advertising Seattle, August 6, 2006 www.sigir2006.org/tutorials.asp

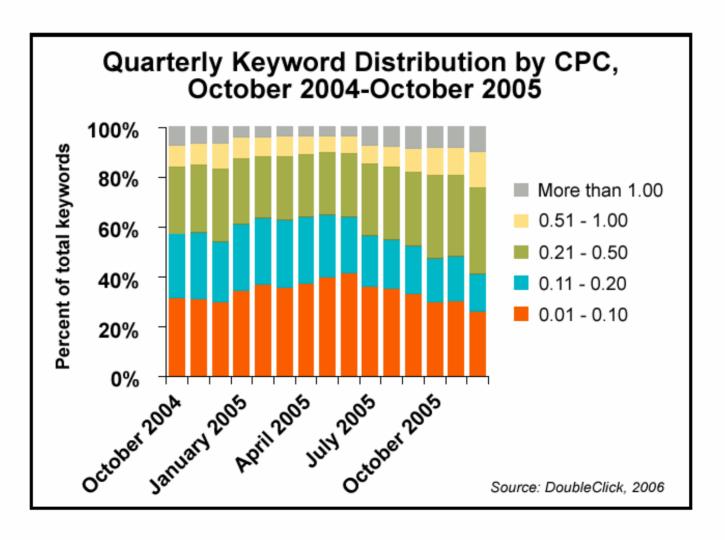
Andrei Broder - Wikipedia, the free encyclopedia

Andrei Broder is a Research Fellow and Vice President of Emerging Search Technology for Yahoo! ... Apoints Dr. Andrei Broder as Research Fellow. Yahoo! ... en wikinedia org/wiki/Andrei Broder - 14k - Cached - Similar nages











Technical challenges

- For advertisers
 - What words to buy?
 - How much to pay?
 - Arbitrage among keywords/suppliers, try to cherry pick demographics
 - Classic optimization problem
 - Spamming is an economic activity ...
- For sellers (search engines)
 - How to price the words?
 - Let the market decide: bidding! (more later)
 - When to place ad? ("Broad match" on non-bidded phrases)
 - How to price "extended" or "broad" match?
 - Forecasting (scale)
 - Long term costs (bad user experience, training people not to click)
- For both
 - What are good words?
 - What extra words to buy?



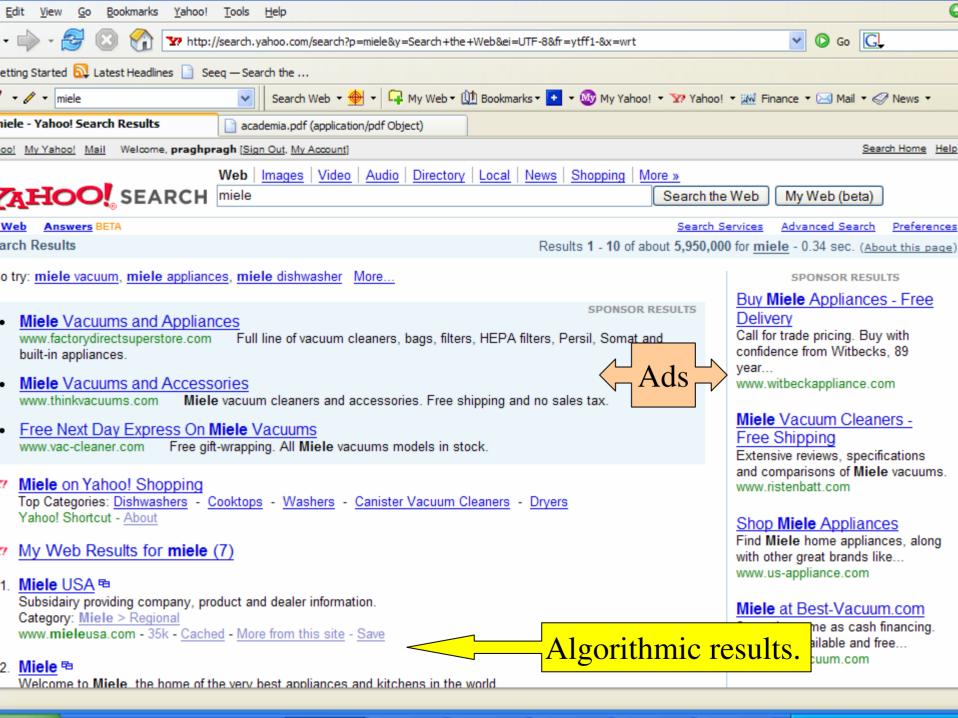
- Relatively simple on "bidded phrases"
- What about queries on which there is no bid?
- Advertiser can bid on "broad queries" and/or "concept queries"
 - Bid on any query that contains "sigir" or the "sigir" concept.
- Pitfalls:
 - A seller of car water pumps might bid on "pumps".
 - How do you avoid "re-built water pumps" ads on "breast pumps"? On "black pumps"?
- Opportunity: Design both algorithms & marketplace!



Finding the right ad (cont)

- Ads Database = Keywords + Title + Description + URL + stats (CTR, bid)
- = Search Keywords + Context Ad Query
- Search problem similar to Web Search, but
 - Ad database entries are "small pages" [+ URL]
 - Different metadata
 - Problem still large (0.5-1B ads)
 - Historical data is enormous (billions of searches/ad/clicks) records)
 - Ranking depends also on bids (more later)

Keyword advertisements: auctions and pricing





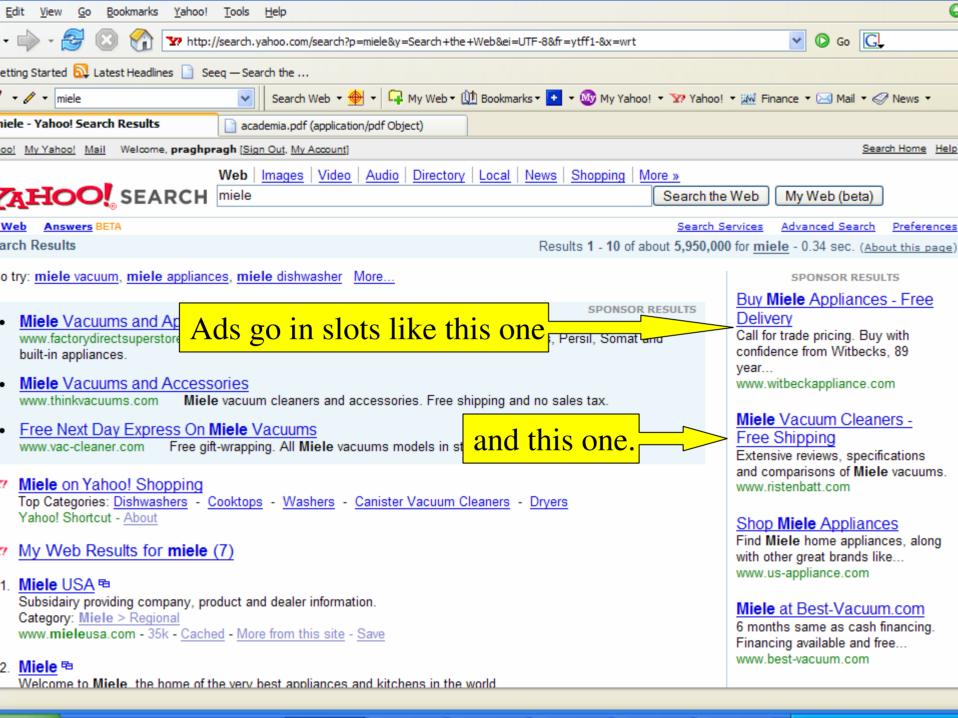
- For each keyword, a *market* in which advertisers compete for attention
 - Advertisers willing to pay for placement on ads next to search results
- Market varies by keyword
 - Commercial vs non-commercial intent in keywords
 - Ford vs Stanford map

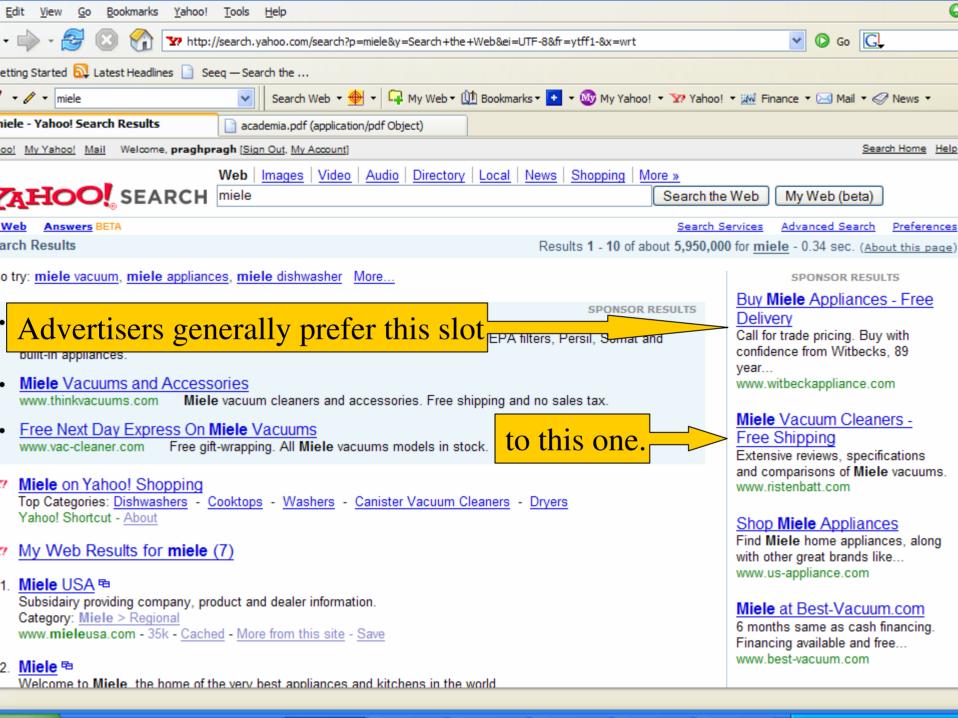
Monetization

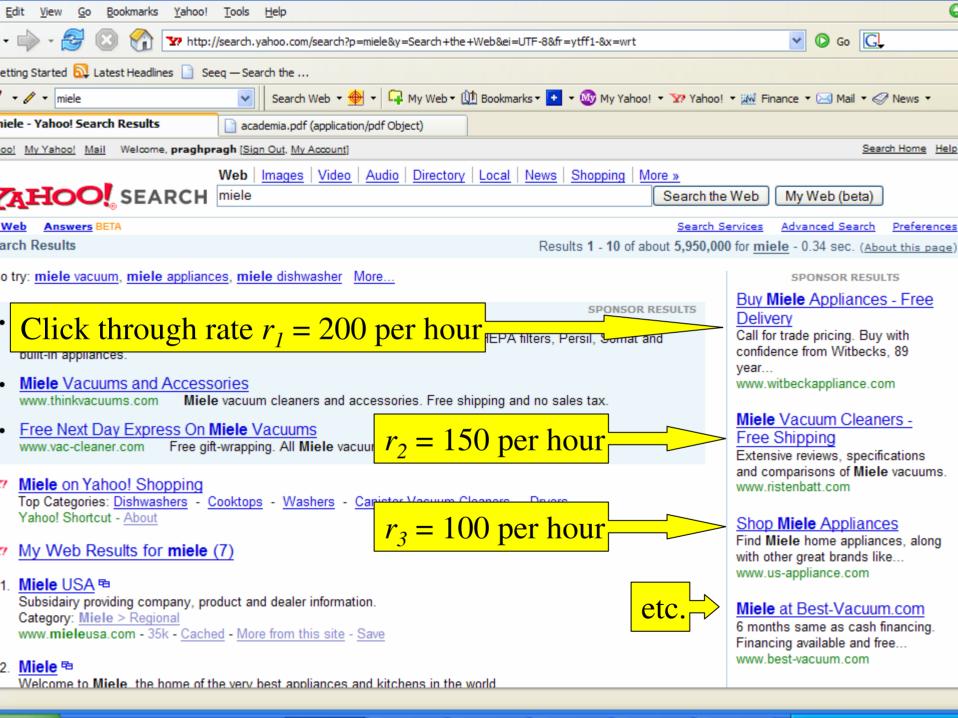
- CPC = When a searcher clicks on a displayed ad, advertiser pays the search engine
 - E.g., a Ford dealership may place an ad for the search *Ford*
 - Out of all such ads, engine chooses to display some on each *Ford* search
 - When a searcher clicks an ad, the advertiser who placed it pays

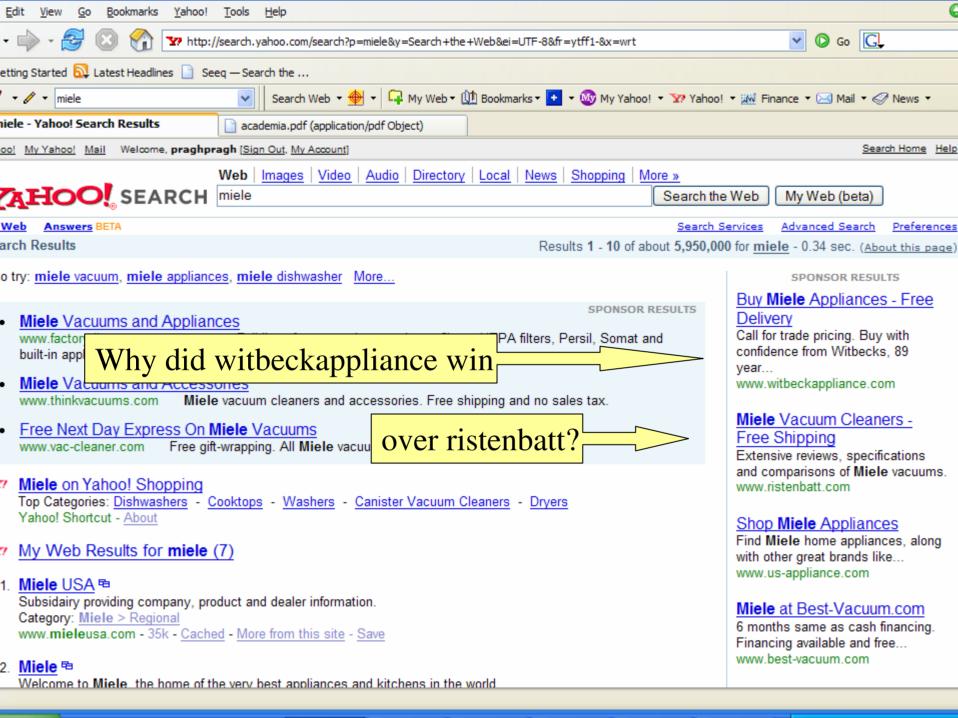


- Of the various advertisers for a keyword, which one(s) get shown?
- What do they pay on a click through?
- The answers turn out to draw from questions in microeconomics









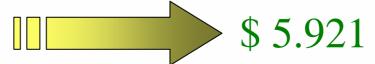


- Click-through rate depends only on the slot, not on the advertisement
- In fact not true; more on this later
 - In reality a function of both the slot and the advertisement.



Advertiser's value

- We'll assume that an advertiser j has a value v_j per click through
 - -Some measure of downstream profit
- Say, user buys a Miele product, e.g.,
 - 96% of the time, no purchase
 - 0.7% buy Dishwasher, profit \$500
 - 1.2% buy Vacuum Cleaner, profit \$200
 - 2.1% buy Cleaning agents, profit \$1





- For the keyword *miele*, an appliance vendor has a value of \$10 per click.
- How much should he bid?
- How much should he be charged?

The <u>value</u> of a slot for an advertiser, what he <u>bids</u> and what he is <u>charged</u>, <u>may all be different</u>.



- We have multiple advertisers, users, and the central service;
- Each has an economic incentive in participation
- How do we set up a system of payments that works for all?
 - Mechanism design
 - Sub-field of game theory



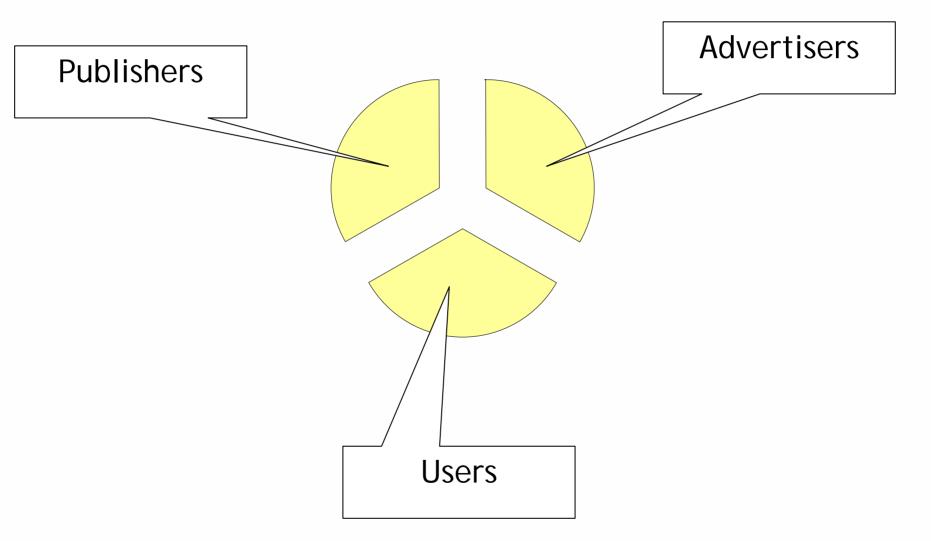
- Study of strategic interactions between two or more <u>rational</u> players
- Classic tool for studying markets
 - Competition as well as collaboration.

Open Questions

- Most real problems are unsolved:
 - Budgets/Allocations
 - Scarcity does not fully explain desire for position 1 – quality of clicks might be different + brand effect
 - These are repeated auctions & there are alternative goods (= synonyms)
 - There are incentives to exhaust the competitors budget.
 - Etc, etc



Contextual ads = meeting of Publishers, Advertisers, Users

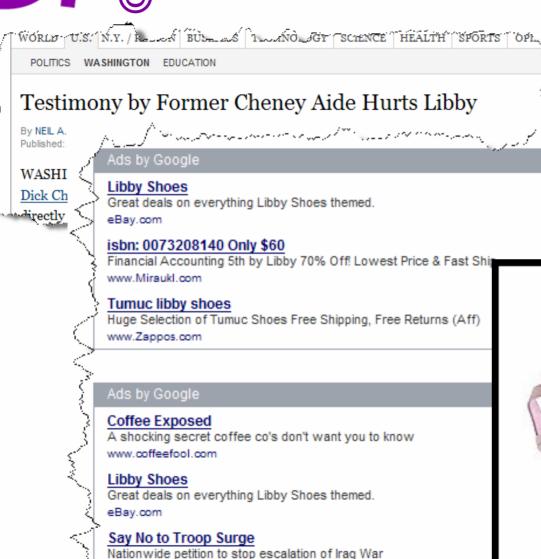




Another problem not quite solved



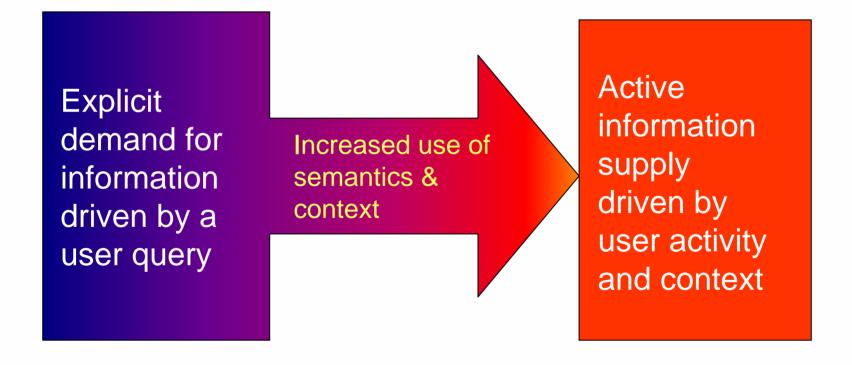
TrueMajority.org





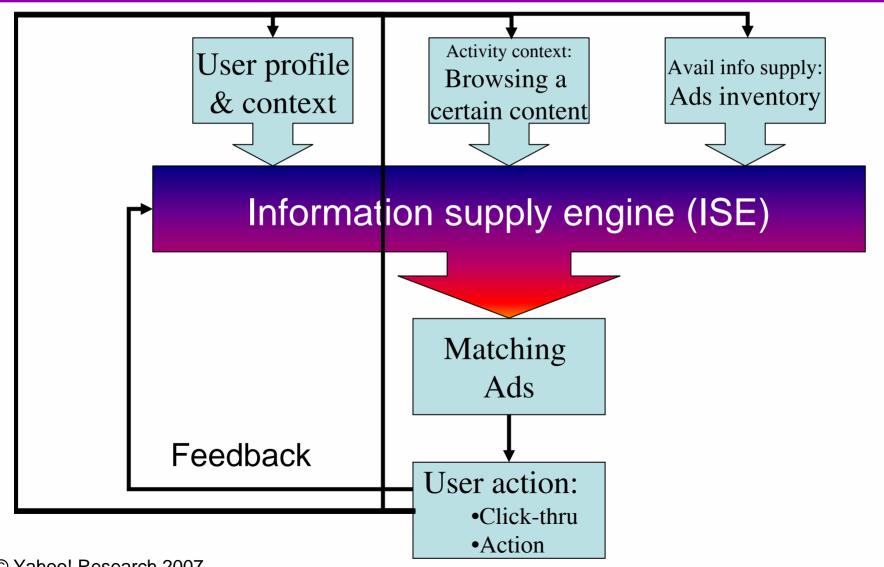


A general trend: From Information Retrieval to Information Supply





Ads as Information supply





Technical challenges to the transition to information supply

- A theory of information supply
- Representation of context
 - Vector model with lots of parameters?
- Representation of information
 - Certainly bag of words is not enough ...
- Representation of user
 - Probabilistic data
- Matching the three above
 - What to use? IR, Huge ML models, HMM, Control Theory, …???



Matching Ads and Content via reduction to search

- Ads Database = Keywords +
 Title + Description + URL
- Ad Query = Landing Page + Context

- Search problem now changes
 - Large query (a page)

Main problem: Semantic Gap



Vocabulary Mismatch

- Synonyms
 - A dictionary partially solves the problem
 - Proper names?
- Spelling problems
 - Approximate match?
- Language mismatch



Business challenges in context driven advertising

- Advertising in general and search driven ads in particular are the economic engine behind the web.
 - To date, contextual ads have been less effective than keyword driven ads
- What is the right pricing model?
- Privacy concerns vs. effectiveness
- Great research opportunities: we can build the technology to solve the business problem, and we can change the business problem to make it solvable.





- Web advertising is scientifically young
- It is intellectually diverse need to deal with:
 - Economics
 - The human element
 - The social element
 - Solving huge problems ridiculously fast
- The technology mirrors the economic, legal and sociological reality



- Fundamental shift:
 - R&D is no longer an upstream function that creates a product then tosses it over to someone else to market, price and sell.
- The economics and marketplace design have to be built into the product design.
 - Else an engineer makes what they think is a technical decision
 - But it turns out to be a marketplace design decision.



Research consequence

- No single discipline has the answer need interdisciplinary approach, involving unusual partners.
- Your chance to learn something new and have a big impact!

Thank you! broder@yahoo-inc.com

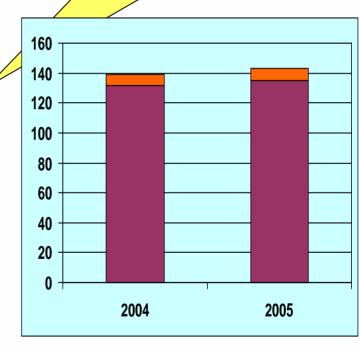


Advertising spend in USA in 2005

Excludes search advertising [TNS Media Intelligence]

Media	Full Year 2005 (Millions)	Full Year 2004 (Millions)	% Change
NEWSPAPERS (LOCAL)	\$25,090.50	\$24,814.40	1.10%
NETWORK TV	\$22,455.20	\$22,523.40	-0.30%
CONSUMER MAGAZINES	\$21,688.00	\$20,167.40	7.50%
CABLE TV	\$15,874.10	\$14,248.80	11.40%
SPOT TV ²	\$15,529.20	\$17,158.70	-9.50%
INTERNET ³	\$8,322.70	\$7,343.00	13.30%
LOCAL RADIO ⁴	\$7,364.90	\$7,273.40	1.30%
B-TO-B MAGAZINES	\$4,471.00	\$4,364.60	2.40%
SYNDICATION - NATIONAL	\$4,222.50	\$3,930.90	7.40%
SPANISH LANGUAGE MEDIA ⁵	\$4,219.20	\$3,976.10	6.10%
OUTDOOR	\$3,528.80	\$3,213.00	9.80%
NATIONAL NEWSPAPERS	\$3,466.70	\$3,303.50	4.90%
NATIONAL SPOT RADIO	\$2,604.10	\$2,616.50	-0.50%
SUNDAY MAGAZINES	\$1,619.50	\$1,497.40	8.20%
FS/s ⁶	\$1,441.50	\$1,391.90	3.60%
NETWORK RADIO	\$1,009.90	\$1,027.80	-1.70%
LOCAL MAGAZINES	\$385.50	\$317.70	21.30%
TOTAL ⁷	\$143,293.40	\$1 39, 1 68.60	3.00%
Sauras: TMS Madia Intelligence			

Internet 2005: \$ 8.3B (+13.3% vs 2004)



Total US 2005: \$ 143.3B

(+3% vs 2004)

Source: TNS Media Intelligence



Search advertising spending

