Technical challenges in web advertising
Andrei Broder
Yahoo! Research
Disclaimer

• 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
60 DAYS OF DAYLIGHT FOR APARTMENT 6F.

SMITHWICK'S IS BRINGING THE LUCK OF THE IRISH TO NYC.
Sometimes both ...
Web Advertising

There are lots of ads on the web ...
Growing Older, and Adjusting to the Dark

By JANE E. BRODY
Published: March 13, 2007

How well do you see at night? If you’re over 50, probably not as well as you think, no matter how many carrots you eat. The typical 50-year-old driver needs twice as much light to see as well after dark as a 30-year-old. Yet few of us compensate adequately for the reduction in nighttime acuity that occurs in the aging eye.
Textual: Sponsored Link:
Text-based ads usually driven by a keyword search
Ad revenue stats from Internet Advertising Board
We concentrate on

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
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)
Key pricing concepts

- **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
CPC Example

- 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
Our new ad in Google

Google

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

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/~ley/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

Andrei Broder - Wikipedia, the free encyclopedia
Andrei Broder is a Research Fellow and Vice President of Emerging Search Technology for
Yahoo! ... Appoints Dr. Andrei Broder as Research Fellow, Yahoo! ... en.wikipedia.org/wiki/Andrei_Broder - 14k - Cached - Similar pages

Work at Google
Google is hiring expert computer scientists and software developers!
www.google.com/jobs

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

© Yahoo! Research 2007
<table>
<thead>
<tr>
<th>Keyword</th>
<th>Status</th>
<th>Current Bid Max CPC</th>
<th>Hide Settings</th>
<th>Clicks</th>
<th>Impr.</th>
<th>CTR</th>
<th>Avg. CPC</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Total</td>
<td>Enabled</td>
<td>Default $0.07 [edit]</td>
<td></td>
<td>44</td>
<td>2,447</td>
<td>1.79%</td>
<td>$0.07</td>
<td>$2.74</td>
</tr>
<tr>
<td>Content Total [?]</td>
<td>Enabled</td>
<td></td>
<td></td>
<td>0</td>
<td>136</td>
<td>0.00%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;www 2006&quot;</td>
<td>Active</td>
<td>$0.07</td>
<td>Settings</td>
<td>26</td>
<td>1,095</td>
<td>2.37%</td>
<td>$0.05</td>
<td>$1.06</td>
</tr>
<tr>
<td>wwww2006</td>
<td>Active</td>
<td>$0.20</td>
<td>Settings</td>
<td>6</td>
<td>509</td>
<td>1.17%</td>
<td>$0.13</td>
<td>$0.78</td>
</tr>
<tr>
<td>&quot;www conference&quot;</td>
<td>Active</td>
<td>$0.10</td>
<td>Settings</td>
<td>9</td>
<td>357</td>
<td>2.52%</td>
<td>$0.09</td>
<td>$0.75</td>
</tr>
<tr>
<td>Prabhakar Raghavan</td>
<td>Active</td>
<td>$0.10</td>
<td>Settings</td>
<td>3</td>
<td>164</td>
<td>1.82%</td>
<td>$0.05</td>
<td>$0.15</td>
</tr>
<tr>
<td>Andrei Broder</td>
<td>Active</td>
<td>$0.07</td>
<td>Settings</td>
<td>0</td>
<td>145</td>
<td>0.00%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CPC costs

Quarterly Keyword Distribution by CPC, October 2004-October 2005

Source: DoubleClick, 2006
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?
Technical issue: Finding the Right Ad

- 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)
- Ad Query = Search Keywords + Context

- 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
A marketplace

• 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
Generic questions

• 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
Ads go in slots like this one and this one.
Advertisers generally prefer this slot to this one.
Click through rate $r_1 = 200$ per hour

$r_2 = 150$ per hour

$r_3 = 100$ per hour

etc.
Why did witbeckappliance win over ristenbatt?
First-cut assumption

• 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

$ 5.921
Example

• 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 value of a slot for an advertiser, what he *bids* and what he is *charged*, may all be different.
Competitive market

• 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
Game theory

- Study of strategic interactions between two or more rational 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
A general trend: From Information Retrieval to Information Supply

Explicit demand for information driven by a user query

Increased use of semantics & context

Active information supply driven by user activity and context
Ads as Information supply

User profile & context

Activity context: Browsing a certain content

Avail info supply: Ads inventory

Information supply engine (ISE)

Matching Ads

User action: • Click-thru • Action

Feedback
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.
Final Remarks
The net

- 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
Change in the game

• 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]

<table>
<thead>
<tr>
<th>Media</th>
<th>Full Year 2005 (Millions)</th>
<th>Full Year 2004 (Millions)</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEWSPAPERS (LOCAL)</td>
<td>$25,090.50</td>
<td>$24,814.40</td>
<td>1.10%</td>
</tr>
<tr>
<td>NETWORK TV</td>
<td>$22,455.20</td>
<td>$22,523.40</td>
<td>-0.30%</td>
</tr>
<tr>
<td>CONSUMER MAGAZINES</td>
<td>$21,688.00</td>
<td>$20,167.40</td>
<td>7.50%</td>
</tr>
<tr>
<td>CABLE TV</td>
<td>$15,874.10</td>
<td>$14,248.80</td>
<td>11.40%</td>
</tr>
<tr>
<td>SPOT TV²</td>
<td>$15,529.20</td>
<td>$17,158.70</td>
<td>-9.50%</td>
</tr>
<tr>
<td>INTERNET³</td>
<td>$8,322.70</td>
<td>$7,343.00</td>
<td>13.30%</td>
</tr>
<tr>
<td>LOCAL RADIO⁴</td>
<td>$7,364.90</td>
<td>$7,273.40</td>
<td>1.30%</td>
</tr>
<tr>
<td>B-TO-B MAGAZINES</td>
<td>$4,471.00</td>
<td>$4,364.60</td>
<td>2.40%</td>
</tr>
<tr>
<td>SYNDICATION - NATIONAL</td>
<td>$4,222.50</td>
<td>$3,930.90</td>
<td>7.40%</td>
</tr>
<tr>
<td>SPANISH LANGUAGE MEDIA⁵</td>
<td>$4,219.20</td>
<td>$3,976.10</td>
<td>6.10%</td>
</tr>
<tr>
<td>OUTDOOR</td>
<td>$3,528.80</td>
<td>$3,213.00</td>
<td>9.80%</td>
</tr>
<tr>
<td>NATIONAL NEWSPAPERS</td>
<td>$3,466.70</td>
<td>$3,303.50</td>
<td>4.90%</td>
</tr>
<tr>
<td>NATIONAL SPOT RADIO</td>
<td>$2,604.10</td>
<td>$2,616.50</td>
<td>-0.50%</td>
</tr>
<tr>
<td>SUNDAY MAGAZINES</td>
<td>$1,619.50</td>
<td>$1,497.40</td>
<td>8.20%</td>
</tr>
<tr>
<td>FSIs⁸</td>
<td>$1,441.50</td>
<td>$1,391.90</td>
<td>3.60%</td>
</tr>
<tr>
<td>NETWORK RADIO</td>
<td>$1,009.90</td>
<td>$1,027.80</td>
<td>-1.70%</td>
</tr>
<tr>
<td>LOCAL MAGAZINES</td>
<td>$385.50</td>
<td>$317.70</td>
<td>21.30%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>$143,293.40</strong></td>
<td><strong>$139,168.60</strong></td>
<td><strong>3.00%</strong></td>
</tr>
</tbody>
</table>

Source: TNS Media Intelligence

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

Total US 2005: $ 143.3B (+3% vs 2004)
Search advertising spending

Click Costs
Spending on search advertising in the U.S.

$10.0 billion

Source: eMarketer