Computational Advertising

Andrei Broder

Yahoo! Research

SCECR, May 30, 2009

Disclaimers

- This talk presents the opinions of the author. It does not necessarily reflect the views of Yahoo! Inc or any other entity
- Algorithms, techniques, features, etc mentioned here might or might not be in use by Yahoo! or any other company

What is “Computational Advertising”?

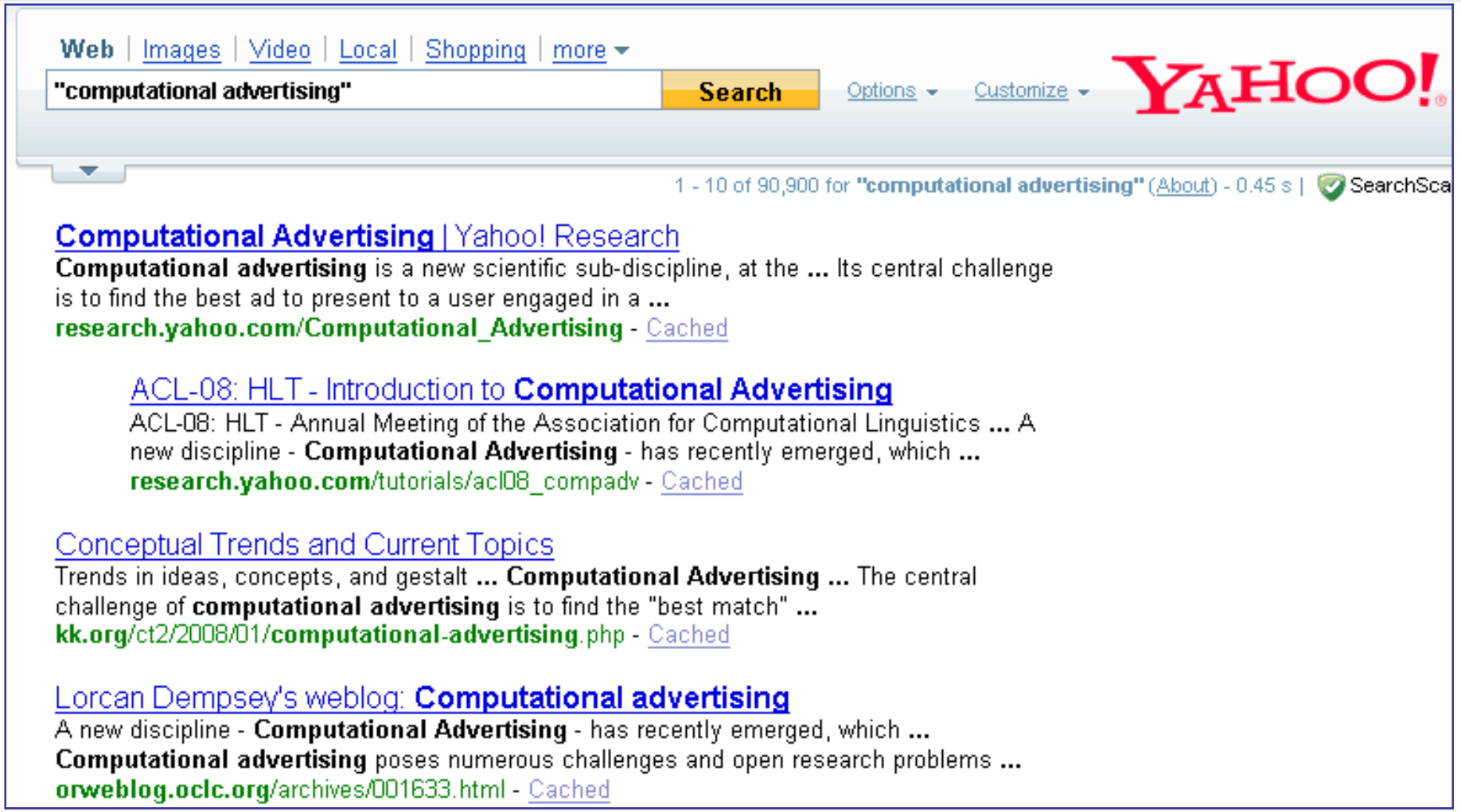
- New scientific sub-discipline, at the intersection of
  - Large scale search and text analysis
  - Information retrieval
  - Statistical modeling
  - Machine learning
  - Classification
  - Optimization
  - Microeconomics
  - Recommender systems

Computational advertising – main challenge

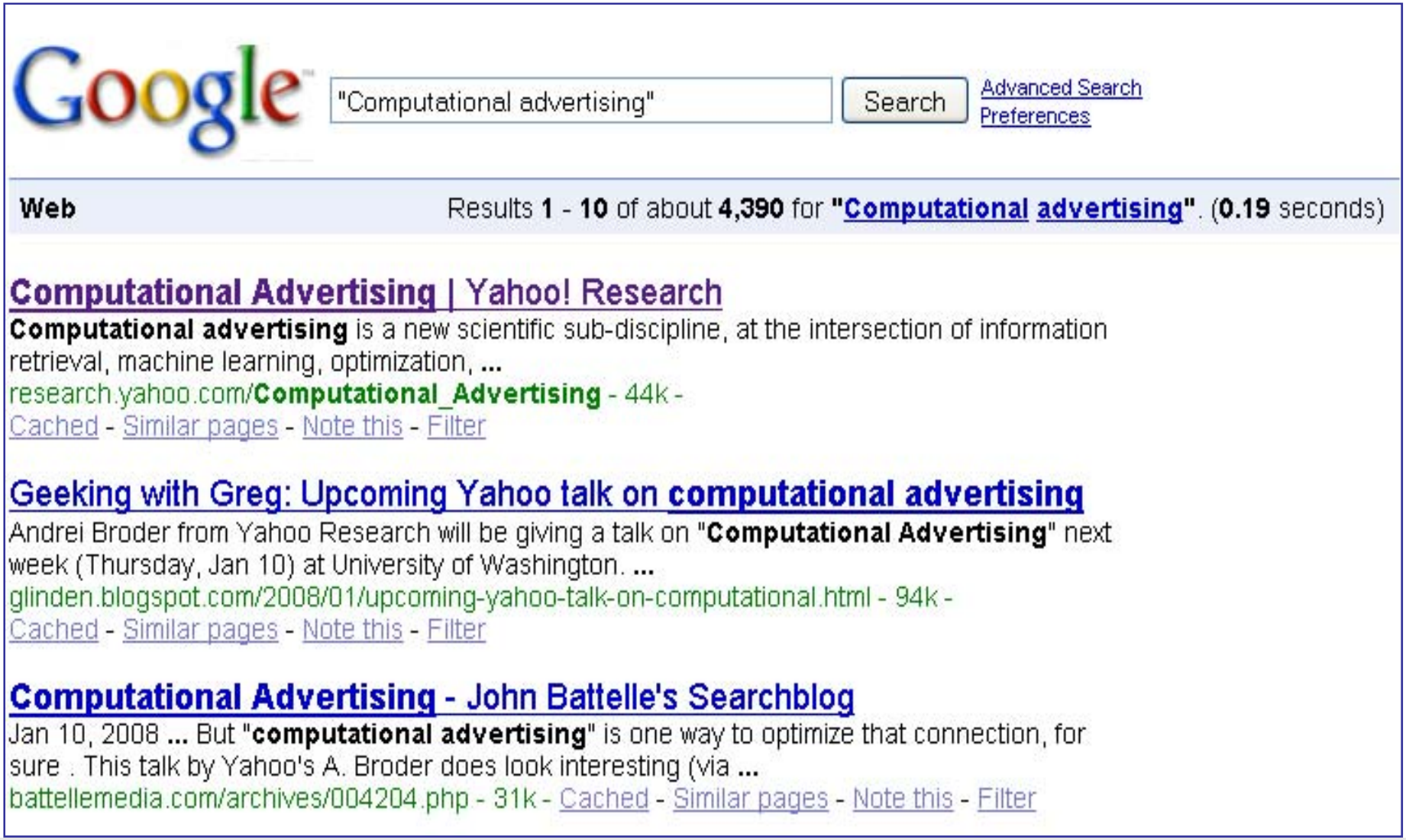
Find the "best match" between a given user in a given context and a suitable advertisement.

- Examples
  - Context = Web search results → Sponsored search
  - Context = Publisher page → Content match, banners
  - Other contexts: mobile, video, newspapers, etc
- What is “best”???

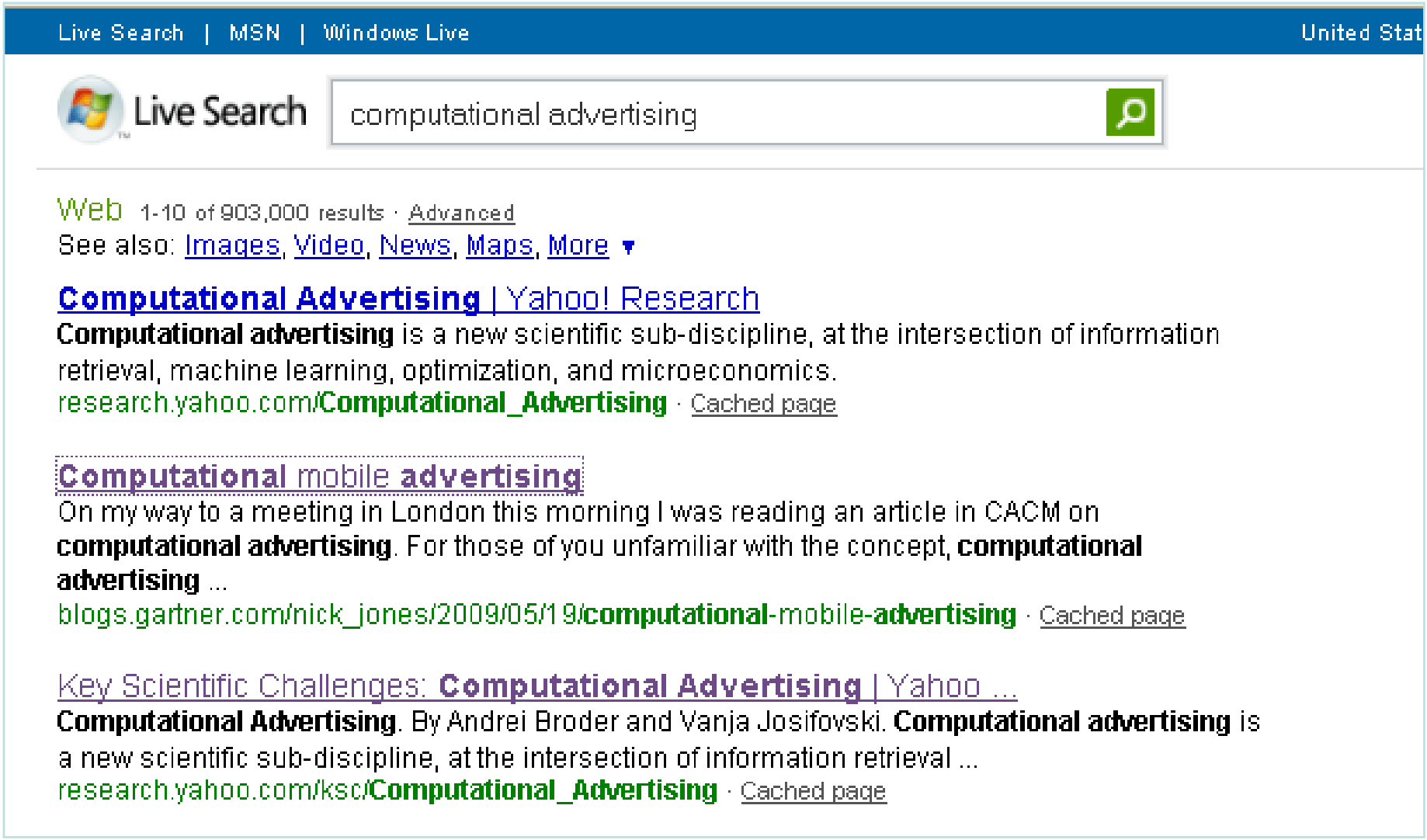
Establishing a new discipline…



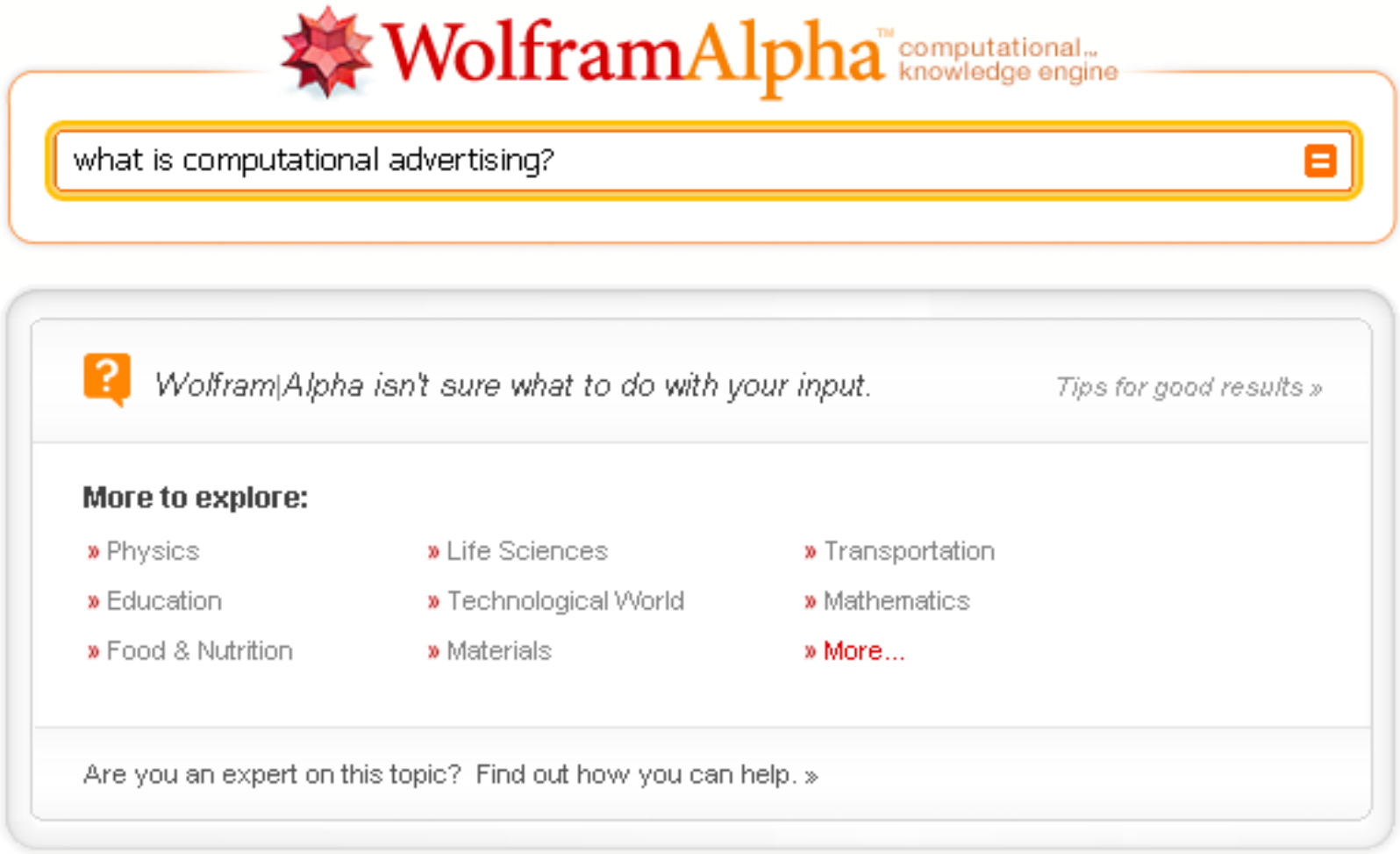
Establishing a new discipline…



Establishing a new discipline…



Establishing a new discipline…



Establishing a new discipline…

Technology | DOI:10.1145/1506409.1506415 David Essex

Matchmaker, Matchmaker

Computational advertising seeks to place the best ad in the best context before the right customer.

Andrei Broder, vice president for computational advertising at Yahoo! Research, presenting a tutorial on Web search and advertising at the 30th Annual International ACM SIGIR Conference in Amsterdam.

THE RAPIDLY CHANGING advertisements that appear on Web pages are often chosen by sophisticated algorithms that match ad keywords to words on a Web page. Take the Chevy ad, for example, that frequently appears on your favorite news site. A real-time ad network at one of the major search engines—Google, MSN, and Yahoo!—might place it on a page of automotive news. But what if the news page's featured article is about a tragic accident caused by a mechanical failure in a Chevy SUV? That's not a page General Motors wants to be associated with, let alone pay good money to advertise on.

Costly mishaps like this could be avoided by a new discipline called computational advertising, which seeks to put the best ad in the best context before the right customer. It draws from numerous fields, including information retrieval, machine learning, natural-language processing, microeconomics, and game theory, and tries to match ads with a variety of user scenarios, such as querying a search engine, reading a Web page, watching a video on YouTube, or instant messaging a friend.

Computational advertising could spur the Web's growth as a medium of mass customization. Better ad matching could quicken the trend toward personalization, making highly specialized magazines, Web sites, and TV channels more financially viable. "Advertising has been the engine that has powered the huge development of the Web," says Andrei Broder, fellow and vice president for computational advertising at Yahoo! Research. "Without advertising, you would not have blogs and search engines."

Computational advertising is a type of automation that tries to replicate what humans might do if they had the time to read Web pages to discern their content and find relevant ads among the millions available. "In the old world of advertising, they deal with few choices and large amounts of money for each choice," Broder says. "We deal with maybe a hundred million potential ads, each worth a fraction of a cent."

A Perfect Match

There are basically three kinds of Web ads. Sponsored search ads are matched to the results of search engine queries; banner ads target particular demographics and venues, typically without regard to a page's content; and contextual advertising, also called context match, applies to other types of Web pages, such as the home page of a financial news site. Computational advertising addresses all three types of ads.

Google, MSN, and Yahoo! use electronic auctions to assign ads to their own results pages and the pages of other Web sites. "Google is a yenta," or matchmaker, says Google chief economist Hal Varian. "The goal is to get a perfect match."

In sponsored search, advertisers bid to place ads that contain keywords correlated to words in a user's search string. For contextual advertising, the keywords are related to words on the entire page, and the search engine's advertising service places the ads. For banner ads, online ad networks place ads on sites whose topics and audiences match the advertiser's criteria.

Before the advent of computational advertising, ad engines could make mistakes more simple-minded than the Chevy SUV scenario. Suppose, for example, a news page contains the word "flowers." If the article isn't about flowers but instead revisits the Rolling Stones' underrated 1967 record *Flowers*, the reader is unlikely to want ads from florists. The old method of analyzing co-occurring words and phrases doesn't help much, and neither does frequency. "You could extract a word used many times in the article and it still is not what the article is about," Broder says.

Therefore, Broder and the 30 researchers who work for him are finding ways to glean the meaning of a page. One promising avenue combines semantic and syntactic features. A semantic phrase categorizes the page and the ads into a 6,000-node topic taxonomy and compares the proximity of the two types of classes as a factor in ranking ads. The hierarchical taxonomy also improves the matching of ads that don't fit a page's exact topic. Keyword matching is still needed to capture more granular content, such as a specific brand of automobile. "We decided that what the article is about should count for about 80% and the words should count for 20%," Broder says.

Another area of interest is using statistical analysis to measure the effect of exogenous events on browsing behavior and adjust the advertisements accordingly. Varian cites short-lived examples, such as this year's rare snowfall in England, or longer-term ones such as the worldwide recession. "In the last few months, there is a big increase in interest in price-sensitive products," Varian says. "The advertisers, in turn, are trying to respond."

All three companies are close-lipped about which of their research has been commercialized, but say that new ideas for algorithms are quickly incorporated into their bidding mechanisms and advertiser tools. Bottom-line results are secret, but the search engines all collect metrics such as revenue per search.

Machine learning, another major focus, concentrates on training algorithms to scan pages for meaning, a technique employed successfully on single-topic documents with the aid of machine-generated labels, but trickier to perform on Web pages, with their assortment of graphics, text, and topics. Microsoft researchers have learned how to employ a type of multiple instance learning to automate classification of sub-documents on pages with incomplete labels and to detect the presence of certain types of content.

"Most of what we do can be boiled down to understanding intent," says Eric Brill, general manager of Microsoft adCenter Labs. By analyzing search strings, for example, algorithms can predict if a person is interested in ads. Some strings are pure attempts at finding information, while others, such as "buy Canon digital camera," have clear commercial intent. "When consumers don't have commercial intent, you don't want to put ads in front of them," Brill says.

Much work focuses on ensuring that new bidding mechanisms don't have incentives for advertisers to misrepresent click-through rates to get better ad placement. In the decentralized economy of the Internet, truthfulness is a currency reinforced by carefully crafted algorithms. "People are out there to make money," says Thore Graepel, a senior researcher at Microsoft Research. "We need to build mechanisms where everyone benefits."

One might expect the speed and volume of data to create a capacity problem, but the researchers express mixed opinions. Graepel says semantic analysis creates an extra burden. "You will hit a computational bottleneck, that's pretty clear," he says. To avoid this, researchers optimize algorithms to make the best decisions with the smallest possible data sets. But they also have faith in engineers' ability to exploit techniques such as parallel processing. "It's surprising how they are always able to scale to deal with these new algorithms," Varian says.

Privacy regulations remain an obstacle to personalizing ads, says Graepel. The existing opt-in, opt-out model lets users choose to reveal personal data in exchange for discounts and other incentives. Researchers are also investigating aggregating data on Web traffic to more accurately match ad categories with coarsely defined groups of users who identify their interests simply by visiting certain types of Web sites.

Fortunately, there is hope for avoiding embarrassments like the ill-placed Chevy ad. Researchers at Microsoft adCenter Labs claim their sub-document classification methods can prevent incompatible ads and Web sites from ever hooking up. You might call it a reverse matchmaker, just the sort of odd little entity the Internet's inventors might never have imagined. ■

David Essex is a freelance science writer based in Peterborough, NH.

Education

Computer Science Enrollment Increases

Enrollment in computer science classes in the United States has increased for the first time in six years, according to the Computing Research Association's (CRA's) annual Taulbee Survey.

Total enrollment by majors and pre-majors in computer science is up 6.2% per department over last year. If only majors are considered, the increase is 8.1%, according to the CRA survey, which collected enrollment data in fall 2008 from computer science and computer engineering departments at 192 Ph.D.-granting universities.

"The upward surge of student interest is real and bigger than anyone expected," says Peter Lee, incoming chair of CRA. "The fact that computer science graduates usually find themselves in high-paying jobs accounts for part of the reversal. Increasingly students also are attracted to the intellectual depth and societal benefits of computing technology."

Computer science graduates on average earn 1.3% more than the average college graduate, according to the U.S. Department of Labor, and future job prospects for computer science graduates are higher than for any other science or engineering field.

The average number of new students per department majoring in computer science is up 9.5% over last year. Computer science departments are replenishing the freshman and sophomore ranks with larger groups than they are graduating as seniors, and computer science graduation rates should increase in two to four years as these new students graduate.

The total number of Ph.D. graduates among responding departments grew to 1,877 for the period July 2007 to June 2008, a 5.7% increase over the previous year.

One area that didn't show improvement is the number of women pursuing computer science degrees, which held steady at 11.8%.

CACM, May 2009: http://mags.acm.org/communications/200905/?pg=18

Key messages

1. Computational advertising = A principled way to find the "best match" between a given user in a given context and a suitable advertisement.
2. The financial scale for computational advertising is huge
   - ⇒ Small constants matter
   - ⇒ Expect plenty of further research
3. Advertising is a form of information.
   - ⇒ Adding ads to a context is similar to the integration problem of other types of information
   - ⇒ Finding the “best ad” is a type of information retrieval problem with multiple, possibly contradictory utility functions
4. New application domains and new techniques are emerging every day
   - ⇒ Good area for research + new businesses

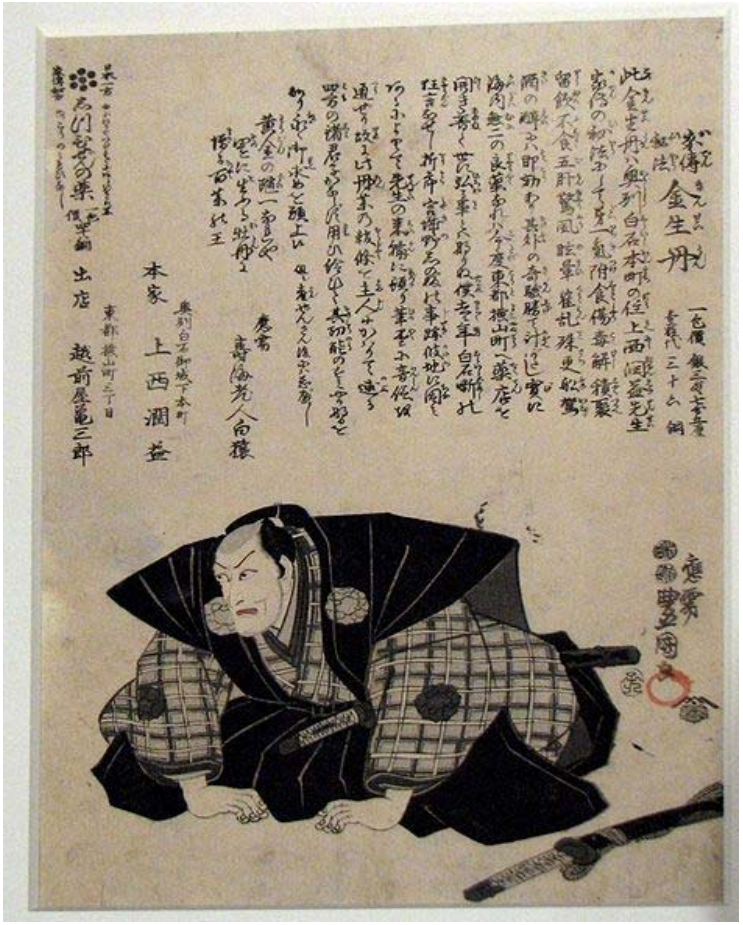
Classic Advertising

Brand advertising
Goal: create a distinct favorable image

Direct marketing

Advertising that involves a "direct response”: buy, subscribe, vote, donate, etc, now or soon

Long history….



Japan ,1806

USA,1890

Lots of computational this and that …

- Computational Biology
- Computational Chemistry
- Computational Finance
- Computational Geometry
- Computational Neuroscience
- Computational Physics
- Computational Mechanics
- Computational Economics
- …

All are about mixing an old science with large scale computing capabilities

What's computational about it?

- **Classical:**
  - Relatively few venues – magazines, billboards, newspapers, handbills, TV, etc
  - High cost per venue ($3 Mil for a Super Bowl TV ad)
  - No personalization possible
  - Targeting by the wisdom of ad-people
  - Hard to measure ROI
- **Computational – almost the exact opposite:**
  - Billions of opportunities
  - Billions of creatives
  - Totally personalizable
  - Tiny cost per opportunity
  - Much more quantifiable

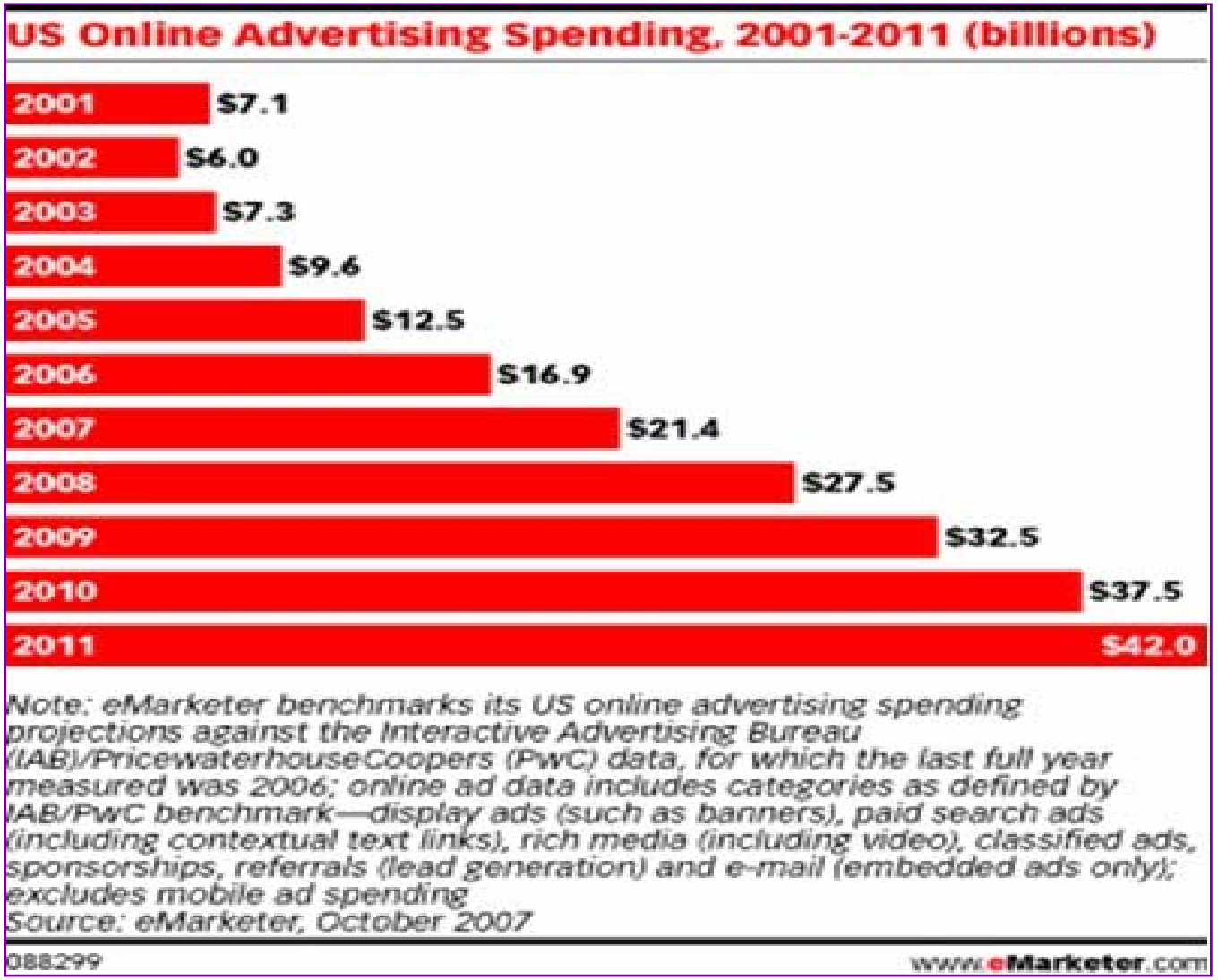
Revenue flow basics

- What do advertisers pay?
  - CPM = cost per thousand impressions
    - Typically used for graphical/banner ads (brand advertising)
    - Could be paid in advance → "Guaranteed delivery"
  - CPC = cost per click
    - Typically used for textual ads
  - CPT/CPA = cost per transaction/action a.k.a. referral fees or affiliate fees
    - Typically used for shopping ("buy from our sponsors"), travel, etc.
    - … but now also used for textual ads (risk mitigation)

- What do publishers get?
  - Whatever advertisers pay minus ***rev-share*** (revenue-share) paid to intermediaries

- What do intermediaries get?
  - Whatever advertisers pay minus ***TAC*** (traffic acquisition costs) paid to publishers

US Online Advertising Spending



Still growing (15% FH 2008 vs. FH 2007)

	FH 2007	FH 2008
Search	41% ($4,097)	44% ($5,064)
Display Related:	32% ($3,198)	33% ($3,799)
-Banner Ads	21% ($2,099)	21% ($2,418)
-Rich Media	7% ($699)	7% ($806)
-Digital Video	1% ($100)	3% ($345)
-Sponsorship	3% ($300)	2% ($230)
Classifieds	17% ($1,699)	14% ($1,611)
Referrals/Lead Generation	8% ($799)	7% ($806)
E-mail	2% ($200)	2% ($230)

Source: IAB

US Online vs. Offline advertising spend

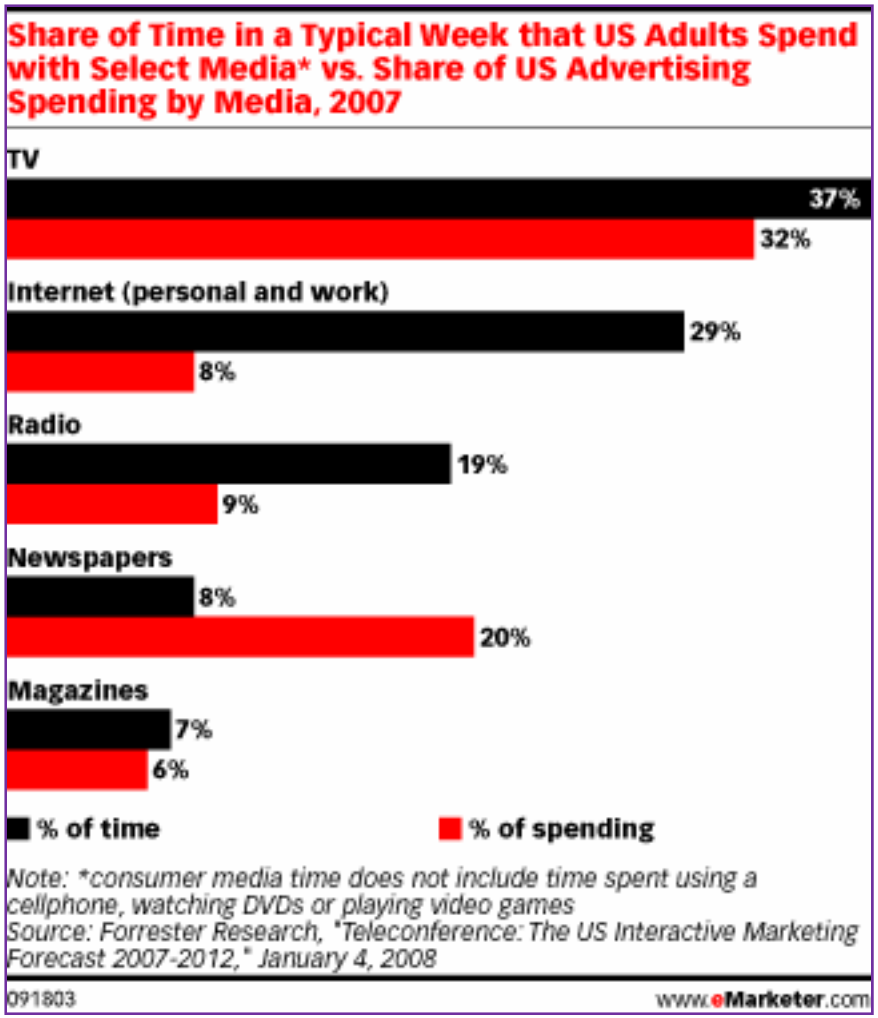
US Online and Total Media Advertising Spending, 2006-2011 (billions and % of total media spending)

	Internet	Total media	Internet % of total media
2006	$16.9	$281.6	6.0%
2007	$21.4	$287.5	7.4%
2008	$27.5	$295.5	9.3%
2009	$32.5	$301.5	10.8%
2010	$37.5	$309.0	12.1%
2011	$42.0	$316.0	13.3%

Note: eMarketer benchmarks its US online advertising spending projections against the Interactive Advertising Bureau (IAB)/PricewaterhouseCoopers (PwC) data, for which the last full year measured was 2006; online ad data includes categories as defined by IAB/PwC benchmark—display ads (such as banners), paid search ads (including contextual text links), rich media (including video), classified ads, sponsorships, referrals (lead generation) and e-mail (embedded ads only); excludes mobile ad spending; eMarketer benchmarks its US total media advertising spending projections against the Universal McCann data, for which the last full year measured was 2006; includes television (broadcast and cable), radio, newspapers, magazines, Internet (excludes mobile), outdoor, direct mail, yellow pages and other
Source: eMarketer, October 2007

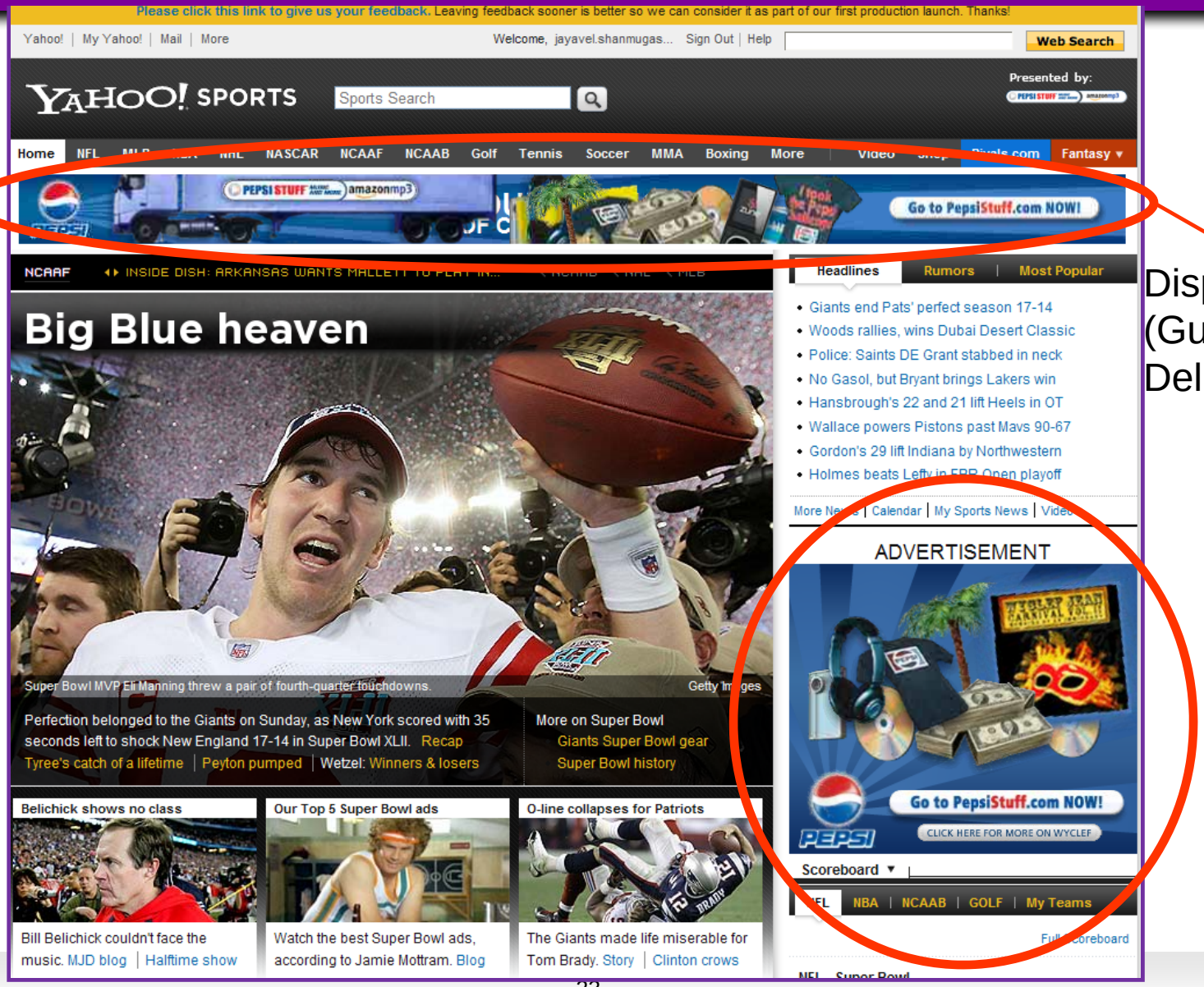
088473 www.eMarketer.com

The advertising $$ budget vs. the human time budget



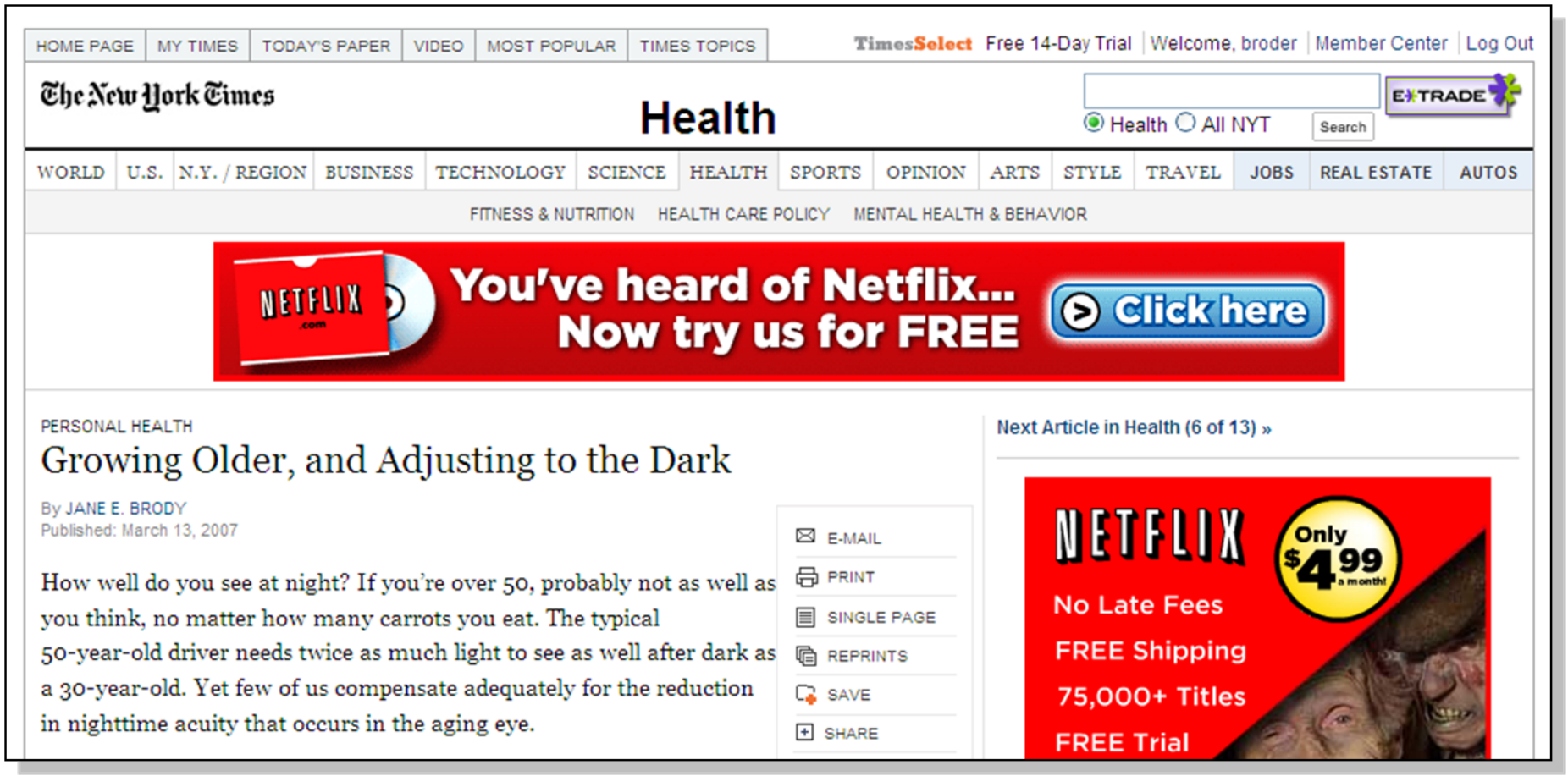
Computational Advertising Landscape

Graphical ads



Display Ads (Guaranteed Delivery)

More graphical ads



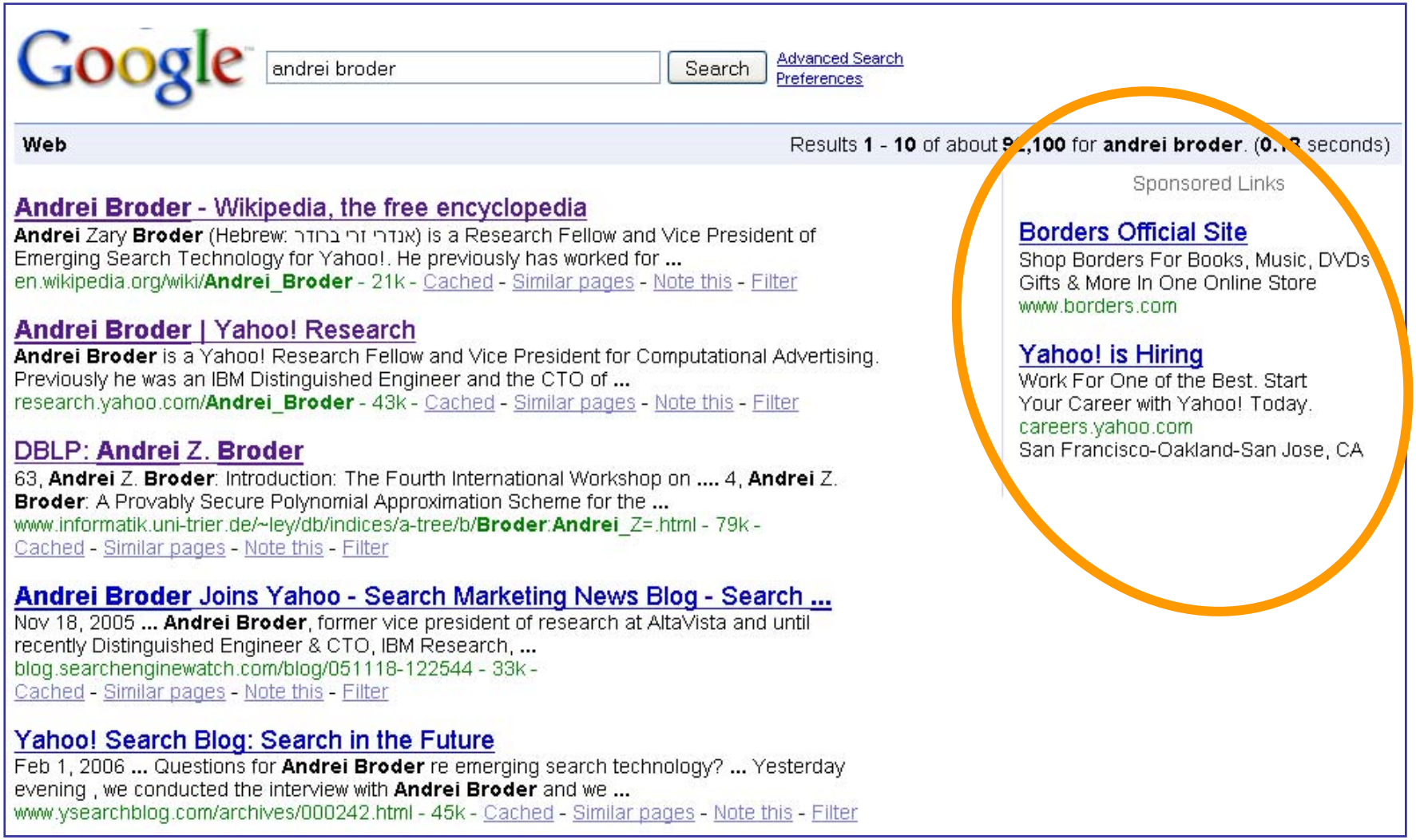
Textual ads

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

Textual ads are heavily related to Search and IR

Sponsored search:

Text-based ads driven by a keyword search



Content match

YAHOO! NEWS

Search

HOME | U.S. | BUSINESS | WORLD | ENTERTAINMENT | SPORTS | TECH | POLITICS | ELECTIONS

Entertainment Video | Celebrity | TV | Movies | Music | Reviews | Fashion | Books | Arts

Beastie Boys recording "political" album

REUTERS

By John Benson – Fri Oct 24, 4:13 am ET

Buzz Up | Send | Share | Print

Reuters – Michael 'Mike D' Diamond (L) and Adam 'MCA' Yauch of the Beastie Boys perform on the main stage during ...

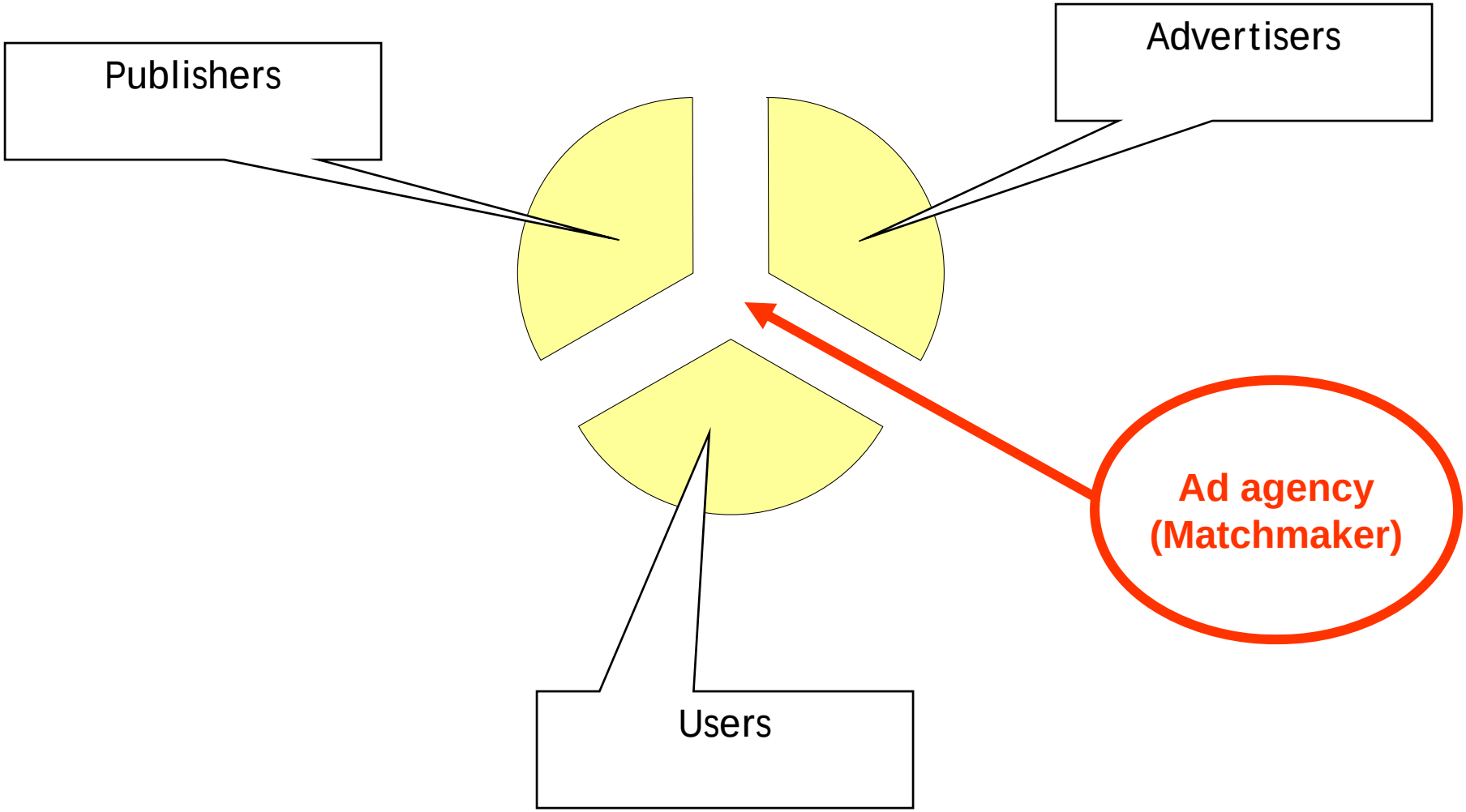
CLEVELAND (Billboard) – A
Get Out and Vote tour, the
2007 instrumental album "T

"We're actually in the middl
Adam "Ad-Rock" Horovitz te
sometime next year. It's a l
And it's political, depending
toilet talk and fart jokes are
yeah, very."

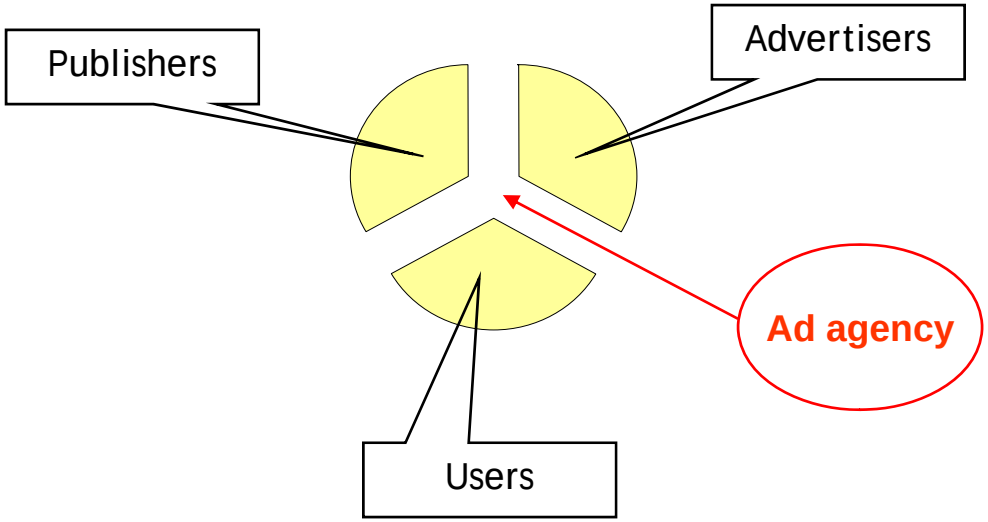
Any chance of new material
08" tour? "I don't think so,"
you play the new songs tha
songs, it always seems like

Video: D.L. Hughley on the election CNN

The actors: Publishers, Advertisers, Users, & “Ad agency”



Dual roles



- Sponsored search:
  - Publisher is also match maker (Yahoo!, Google)
- Content match:
  - Publisher is also match maker (Yahoo! content)
  - Publisher is also advertiser (“House Ads”)

Advertising as information

- “I do not regard advertising as entertainment or an art form, but as a medium of information….” [David Ogilvy, 1985]
- “Advertising as Information” [Nelson, 1974]
- Irrelevant ads are annoying; relevant ads are interesting
  - Vogue, Skiing, etc are mostly ads and advertorials

Finding the “best ad” as an Information Retrieval (IR) problem

1. Analyze the “query” and extract query-features
   – Query = full context (content, user, environment, etc)
2. Analyze the documents (= ads) and extract doc-features
3. Devise a scoring function = predicates on q-features and d-features + weights
4. Build a search engine that produces quickly the ads that maximize the scoring function

Behind the scenes…

- **Setting the ad retrieval problem:**
  - **Ads corpus =**
    - Bid phrase(s) + Title + Creative + URL + Landing Page + …
  - **Query features =**
    - Search Keywords + Outside Knowledge Expansion + Context features
  - **Context features (for sponsored search) =**
    - Location + User data + Previous searches + …
  - **Context features (for context match) =**
    - Location + User data + Page topic + Page keywords …

- Search problem similar to web search, but
  - Ad database is smaller
  - Ad database entries are “small pages” [+ URL]
  - Ranking depends also on bids
  - Ranking depends also on click-through-rate

What is the best match?

- An ad has different utilities for publishers, advertisers, users
- Quality (utility) Factor (QF) is different
  - A-QF, U-QF, P-QF
- The ad agency has its own economic interest
- Might have different types of ads that are not easily compared
- Might have economics/contractual obligations that need to be fulfilled.

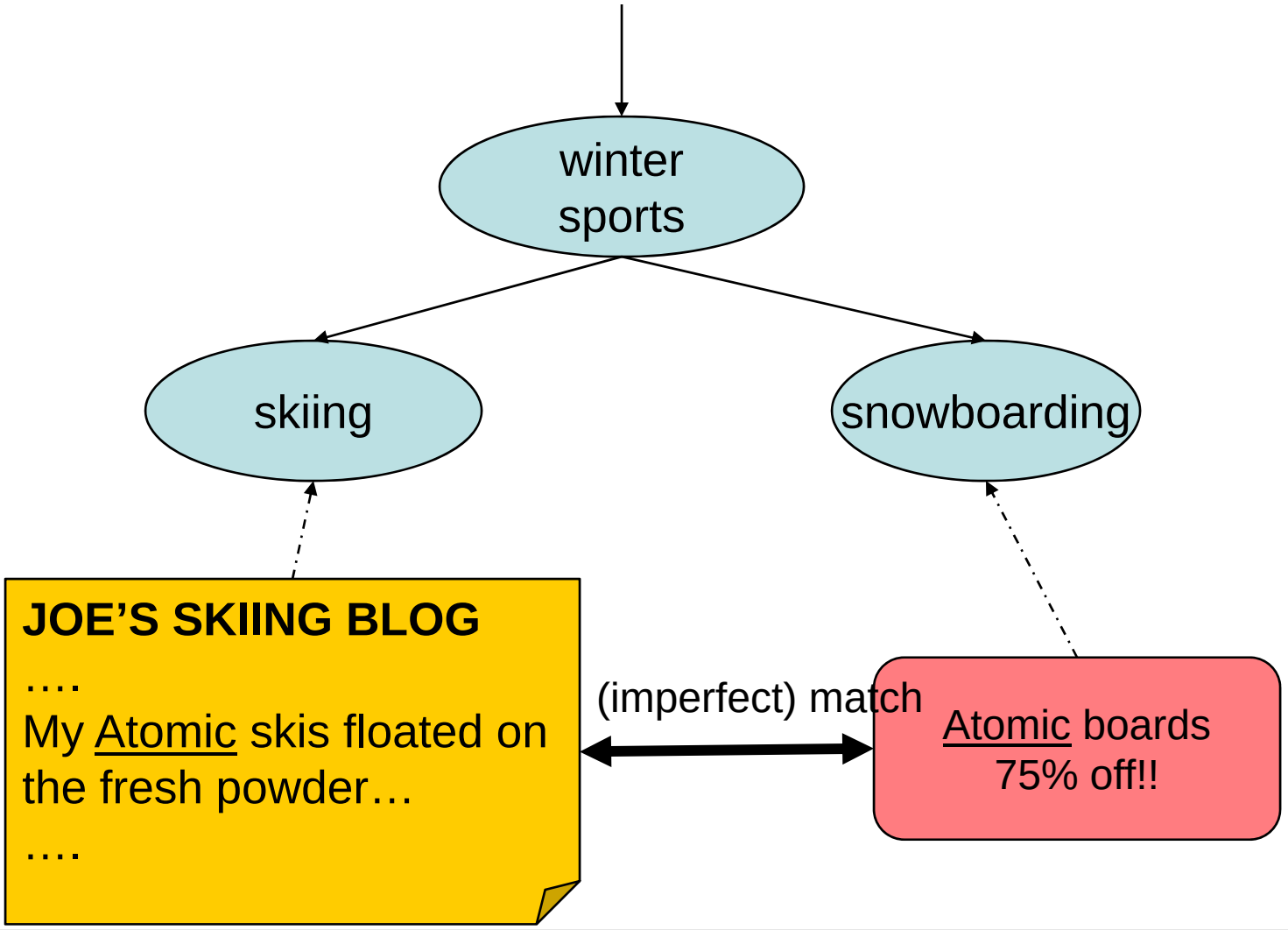
A bit deeper: how does the matching happens?

A Semantic Approach to Contextual Advertising [SIGIR 2007]

[AB, M. Fontoura, V. Josifovski, & L. Riedel]

- What is more important: the words or the context?
- Contextual ad matching based on a combination of semantic and syntactic features.
- Classify both ads and pages into a 6000 nodes commercial taxonomy
- The class information captures the “about-ness” of pages and ads

Matching ads and pages



Semantic and syntactic score

Semantic component – weighted taxonomy distance

$$Tax(d_j) = \{d_{j1} \dots d_{jv}\} \qquad \sum_{d \in Tax(x_i)} cWeight(d) = 1 \qquad idist(c,p) = \frac{n_c}{n_p}$$

$$TaxScore(PC, AC) = \sum_{pc \in PC} \sum_{ac \in AC} idist(LCA(pc, ac), ac) \cdot cWeight(pc) \cdot cWeight(ac)$$

Syntactic component - term vector cosine

$$tWeight(kw^{si}) = weightSection(S_i) \cdot tf\_idf(kw)$$

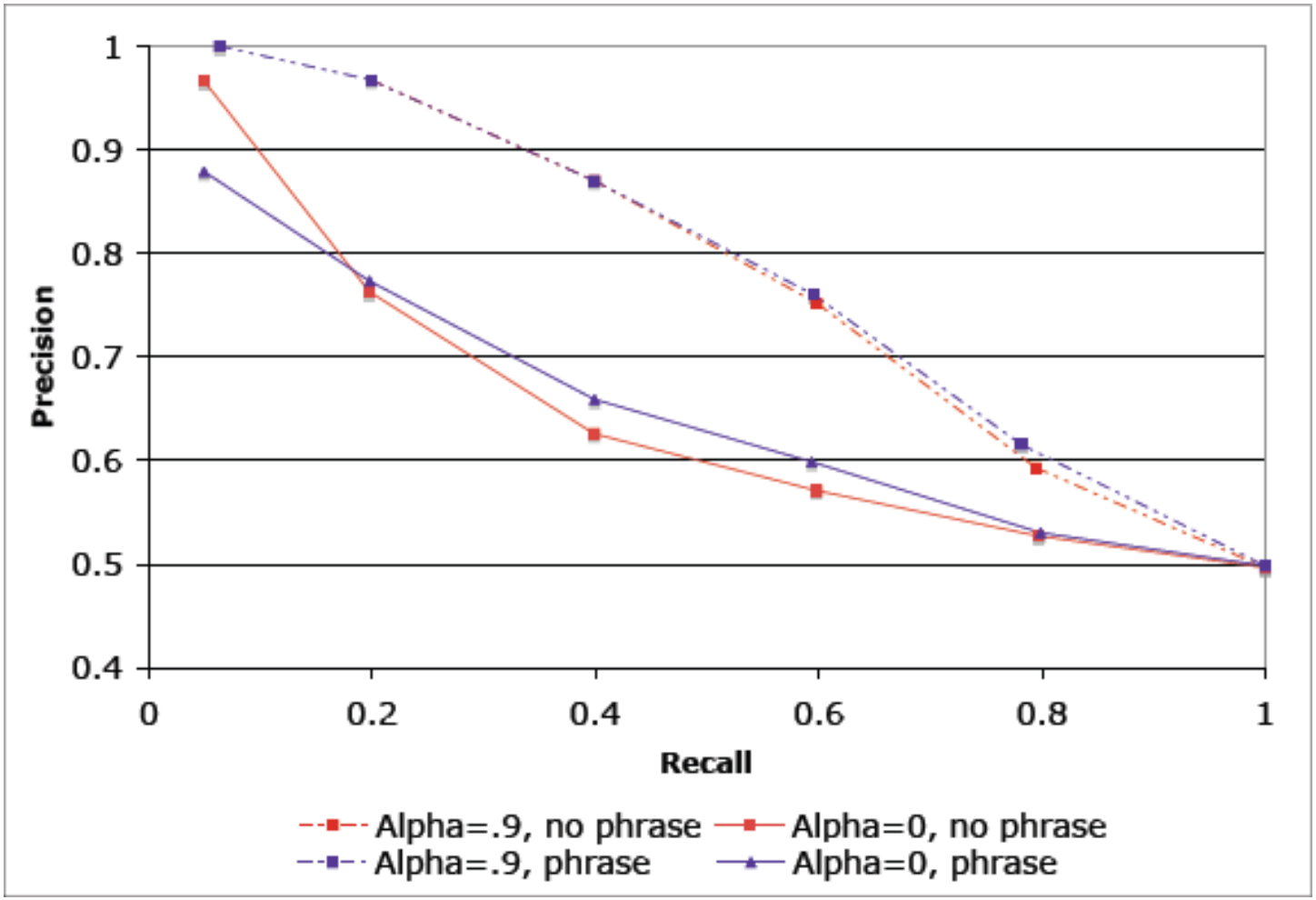
$$KeyordScore(p_i, a_i) = \frac{\sum_{i \in |K|} tWeight(pw_i) \cdot tWeight(kw_i)}{\sqrt{\sum_{i \in |K|} (tWeight(pw_i))^2} \sqrt{\sum_{i \in |K|} (tWeight(aw_i))^2}}$$

Final score

$$Score(p_i, a_i) = \alpha \cdot TaxScore(Tax(p_i), Tax(a_i))$$

$$+(1 - \alpha) \cdot KeywordScore(p_i, a_i)$$

Results


1
0.9
0.8
0.7
0.6
0.5
0.4
Precision
0
0.2
0.4
0.6
0.8
1
Recall
Alpha=.9, no phrase
Alpha=0, no phrase
Alpha=.9, phrase
Alpha=0, phrase

The recommender systems connection

The ad matching problem as a recommendation challenge

- The traditional IR approach is based on a fixed query ←→ results correspondence
- For ads we
  - Need CTR probability or user utility rather than top-K results
  - Have a continuous click-through feedback
- Challenge: incorporate the feedback
  1. Long term loop: improve the ranking function
     - ML based ranking
  2. Short term loop: use the statistics we have for a particular (query, ads) pair
     - Closest to recommender systems

A generic view – dyadic interaction systems

- D. Agarwal, B.C. Chen “Feature based factorization model for dyadic data” [In preparation]
- Dyadic Interaction data pervasive
  - Recommendation systems (user-movie, user-music, user-book)
  - Web advertising (match ads to webpage/query)
  - Content Optimization (match articles to users)
- Unit of measurement : dyad →(i, j)
  - i= user, webpage,..; j= movies, ads,…
  - Measure some response: ratings, click-rates,…
  - Often have meta-data on dyadic elements
    - Demographics, genres,….
  - Goal: predict response for unknown dyads
    - Better match-making, prediction

Challenge: clicks are very rare

- Need to aggregate clicks/recommendations
  - Pages belong to site section to sites
    Page $\in$ CNN/sports $\in$ CNN
  - Ads belong to Campaign to Advertisers
    Ad $\in$ Ford Focus $\in$ Ford
  - Similar to preference estimation for
    Annie Hall $\in$ Woody Allen Comedies $\in$ Comedies
- D. Agarwal, A. B, D. Chakrabarti, D. Diklic, V. J., and M. Sayyadian. Estimating rates of rare events at multiple resolutions. [KDD, 2007]

Challenge: how to find new ads?

- Similar to music recommender systems: need to explore new songs but need to keep some similarity
- Can take advantage of semantic taxonomy
  - S. Pandey, D. Agarwal, D. Chakrabarti, and V. Josifovski. Bandits for taxonomies: A model-based approach. [SDM, 2007]

Summary

Key messages

1. Computational advertising = A principled way to find the "best match" between a given user in a given context and a suitable advertisement.
2. The financial scale for computational advertising is huge
   - ⇒ Small constants matter
   - ⇒ Expect plenty of further research
3. Advertising is a form of information.
   - ⇒ Adding ads to a context is similar to the integration problem of other types of information
   - ⇒ Finding the “best ad” is a type of information retrieval problem with multiple, possibly contradictory utility functions
4. New application domains and new techniques are emerging every day
   - ⇒ Good area for research + new businesses

Thank you!

broder@yahoo-inc.com

http://research.yahoo.com