Introduction to Computational Advertising

MS&E 239
Stanford University
Autumn 2011
Instructors: Dr. Andrei Broder and Dr. Vanja Josifovski, Yahoo! Research
General course info

- Course Website: [http://www.stanford.edu/class/msande239/](http://www.stanford.edu/class/msande239/)
- TA: Krishnamurthy Iyer
  - Office hours: Tuesdays 6:30pm-7:00pm, Huang
- Course email lists
  - Staff: [msande239-aut1112-staff](mailto:msande239-aut1112-staff)
  - All: [msande239-aut1112-students](mailto:msande239-aut1112-students)
  - Please use the staff list to communicate with the staff
- Lectures: 10am ~ 12:30pm Fridays in HP
- Office Hours:
  - After class
  - Andrei and Vanja will be on campus for 2 times each to meet and discuss with students. Feel free to come and chat about even issues that go beyond the class.
Instructor

- Dr. Vanja Josifovski
  - Senior Director at Yahoo! Research
  - Research Area: Computational Advertising, Search
  - Previously at IBM Research working on databases and enterprise search
  - M.Sc. from University of Florida, PhD from Linkopings University in Sweden
    - vanjaj@yahoo-inc.com
    - http://research.yahoo.com/Vanja_Josifovski
Instructor

• **Dr. Andrei Broder**
  • Fellow and Vice President for Computational Advertising in Yahoo! Research
  • Chief Scientist of Yahoo’s Advertising Technology Group
  • Research interests: computational advertising, web search, context-driven information supply, and randomized algorithms
  • B. Sc. Summa cum Laude from the Technion, M.Sc. and Ph.D. in Computer Science at Stanford University under Don Knuth
    • broder@yahoo-inc.com
    • http://research.yahoo.com/Andrei_Broder
Course Overview (subject to change)

1. 09/30 Overview and Introduction
2. 10/07 Marketplace and Economics
3. 10/14 Textual Advertising 1: Sponsored Search
4. 10/21 Textual Advertising 2: Contextual Advertising
5. 10/28 Display Advertising 1
6. 11/04 Display Advertising 2
7. 11/11 Targeting
8. 11/18 Recommender Systems
9. 12/02 Mobile, Video and other Emerging Formats
10. 12/09 Project Presentations
This lecture

1. Logistics of the course
2. Introduction & Overview of Computational Advertising
3. Detailed discussion of the requirements, goals of the course, questions, anything else
Logistics
General lecture structure

- Overview of the area: 1:15 hour
- Break: 10 minutes
- Student presentation 15 min
- In depth, discussion, and occasional quizzes: 1 hour
Class requirements

- Homework 40%
- Project 40%
- Quizzes 10%
- In-class short presentation 10%
- Active attendance strongly encouraged!
Grading and Requirements
Projects – 40%

- Group assignment
  - Preferred group size is 3
- Start forming groups!
- Project commenced at the second class
- Chose one of two projects
  - Real world online advertising
  - Algorithmic project
Real world advertising project

- Craft a campaign for a real world company
- Find a company that can benefit from online advertising
  - Local business
  - Upcoming company (startup)
- Funding provided ($300)
- Craft campaigns
  - Major search engines
  - Display advertising Facebook/Google/Yahoo/MSFT
  - Smaller outlets that allow for customization/optimization
- Adjust campaign mid flight
- Important to learn something insightful!
- Write report and present at the last lecture (15min)
Algorithmic project

- Implement a small scale similarity search engine
- We provide a dataset based on a real world ad corpus
- Build an inverted index
- Implement two algorithms
- Explore multiple variants of the algorithms
- Write a report and present at the last class (15min)
Homework – 40%

- Total of 3 homework assignments
- Take home, open book one week to finish
- Third assignment to be a take-home, final with problems covering the while course
- Homework are individual assignments
  - Please do not work together on the homework assignments
  - If you have any questions please contact the staff
Quizzes – 10%

- Total of 2 quizzes
- In class after the break for 10 minutes
- A few short questions covering the salient points from the class
- The goal of the quizzes is to motivate attendance
In-class presentation – 10%

- Every student should participate in a in-class presentation
- after the break, 15 minutes
- Describe the business model and the technology of a given company in the computational advertising field
- Kris, our TA will do a presentation in the next class to demonstrate the format
- Please start forming groups (3-4 people per group), and contact Kris about the scheduling.
- We might need to have 2 presentations in some classes to accommodate everybody.
Introduction to Computational Advertising
Disclaimers

- This talk presents the opinions of the authors. 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.
- These lectures benefitted from the contributions of many colleagues and co-authors at Yahoo! and elsewhere. Their help is gratefully acknowledged.
Lecture 1 plan

- Overview and key messages
- Classical advertising
  - Difference between classic and computational
- The opportunity: revenue and beyond
- Computational Advertising landscape
  - Graphical ads: guaranteed delivery, performance delivery, exchanges
  - Textual ads
- Ad selection
  - Textual ad selection
  - Performance graphical ads
- Mobile advertising
- Closing remarks
Computational advertising – the central 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
Participants: Publishers, Advertisers, Users, & “Matcher”
What is “Computational Advertising”? 

- New scientific sub-discipline bringing together
  - Information retrieval
  - Large scale search and text analysis
  - Statistical modeling
  - Machine learning
  - Microeconomics
  - Game theory, auction theory, mechanism design
  - Classification
  - Optimization
  - Recommender systems
  - ....
Establishing a new discipline…

**Computational Advertising** | **Yahoo! Research**
Computational advertising is a new scientific sub-discipline, at the ... Introduction to Computational Advertising, Course at Management Sciences and ... research.yahoo.com/Computational_Advertising - Cached

**Phil Windley's Technometria | Computational Advertising**
Organizations Get the ... Computational Advertising. Andrei Broder of Yahoo! Research ... Computation Advertising on, what else, computational advertising ... windley.com/archives/2008/04/computational_advertising.shtml - Cached

**Computational advertising | Lorcan Dempsey's weblog**
A new discipline - Computational Advertising - has recently emerged, which ... Computational advertising poses numerous challenges and open research problems ... orweblog.oclc.org/archives/001633.html - Cached

**Computational Advertising**
Trends in ideas, concepts, and gestalt ... The central challenge of computational advertising is to find the "best match" ... Personalized advertising is a ... kk.org/ct2/2008/01/computational-advertising.php - Cached

**Key Scientific Challenges: Computational Advertising | Yahoo! ...**
Computational Advertising. By By Andrei Broder and Vanja Josifovski. Computational advertising is a new scientific sub-discipline, at the ...
Establishing a new discipline…

computational advertising

About 955,000 results (0.18 seconds)

Computational Advertising | Yahoo! Research
Computational advertising is a new scientific sub-discipline, at the intersection of information retrieval, machine learning, optimization, ...
research.yahoo.com/Computational_Advertising - Cached - Similar - Filter

Stanford University - Introduction to Computational Advertising
14 Nov 2009 ... Computational advertising is a new scientific discipline, at the intersection of information retrieval, machine learning, optimization, ...
Course Information - Course Schedule - Lecture Handouts
www.stanford.edu/class/msande239/ - Cached - Similar - Filter

Geeking with Greg: Lectures on Computational Advertising
4 Jan 2010 ... Slides from all the lectures of Andrei Broder's recent Computational Advertising class at Stanford University now are available online ...
glinden.blogspot.com/.../lectures-on-computational-advertising.html - Cached - Similar - Filter

[Introduction to computational advertising]
File Format: PDF/Adobe Acrobat - Quick View
From IR to IS. ■ Advertising on the Web ...

Computational advertising
by A Broder - 2008 - Cited by 3 - Related articles
Establishing a new discipline…

Computational Advertising | Yahoo! Research

Computational advertising is a new scientific sub-discipline, at the intersection of information retrieval, machine learning, optimization, and microeconomics.
research.yahoo.com/Computational_Advertising · Cached page

Computational advertising and recommender systems
ACM: Association for Computing Machinery SIGKDD: ACM Special Interest Group on Knowledge Discovery in Data
doi.acm.org/10.1145/1454008.1454009 · Cached page

Computational Advertising - John Battelle's Searchblog
I think humans are required anytime you want to connect a brand with a person in any kind of meaningful way. But "computational advertising" is one way to optimize that connection ...
battellemedia.com/archives/004204.php · Cached page

Phil Windley's Technometria | Computational Advertising
Organizations Get the IT They Deserve … Phil tweeted, "Registering my son, who's in Brazil, for classes at BYU. Fun times at midnight."
www.windley.com/archives/2008/04/computational_advertising.shtml · Cached page
Technology | 00.13.14/15:1589421.1509413

Matchmaker, Matchmaker
Computational advertising seeks to place the best ad in the best context for the right consumer.

T
he rapidly changing advertising methods that appear on Web pages are often chosen by sophisticated algorithms that match an ad’s 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 target 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 general Motors warranty situation but one associated with tire blowouts and bad weather, and it’s likely to pay good money to advertise on it.

Costly missteps like this could be avoided by a new discipline called computational advertising, which seeks to match the best ad to the best context for the right consumer in real time. In a number of fields, including information retrieval, machine translation, spam filtering, microeconomics, and game theory, researchers understand how to algorithmically generate scenarios with a high probability 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 quickly cement the trend toward personalized, highly specific advertising. Web sites, and even e-commerce customers, will be more financially viable. Advertising has been the engine that has powered the huge development of the Web,” says Andre 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 rates how relevant an ad is to what consumers actually search for and click on. Computational advertising engines then bid on ads in order to place them on the ad site in a way that is as relevant as possible.

Machine learning, another major focus, concentrates on training algorithms to learn from complex data sets and develop a technique employed successfully on single-topic documents with the aid of machine-generated labels, but this fails to perform on Web pages, with their mass amounts of text on graphics, text, and video.

Microsoft researchers have learned how to apply a type of multiple-instance learning to automate categorization 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. Home settings are pure arenas at finding information, while others, such as "Sony Digital cameras,” have clear commercial intent. "When consumers don’t have commercial intent, you don’t want those ads in front of them,” Brill says.

Much work focuses on ensuring that new bidding mechanisms don’t have incentives for advertisers to mispresent click-through rates to get better ad placement. In the decentralized econo-

Education
Computer Science Enrollment Increases

Enrollment in computer science classes at the University of Texas has increased by 320% in the past five years, according to the Computer Science undergraduate Associations (CSUAs) annual End of Term Survey.

Total enrollment by majors and pre-engineer computer science is 30.6% higher per university. Of these students, 28.6% are women.

The average number of first-year students per department decreased in the past four years, according to the new survey.

The average number of first-year students per department declined in the past four years, according to the new survey.

One oft-cited concern about the high percentage of women pursuing computer science degrees is the lack of support and retention. For example, the percentage of women in computer science degrees declined, which helps to solidify a 15.6% decrease in the past four years to 59.6%.

In contrast, the percentage of women in computer science degrees declined, which supports the idea that women are more likely to pursue computer science degrees. For example, the percentage of women in computer science degrees declined, which helps to solidify a 15.6% decrease in the past four years to 59.6%.

The total number of female students who graduated increased by 32.7% in the past four years, according to the new survey. The percentage of women pursuing computer science degrees increased, which helps to solidify a 15.6% decrease in the past four years to 59.6%.
Key messages

1. **Computational advertising** = A principled way to find the "best match" between a user in a context and a suitable ad.

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
Long history: Advertising for coffee, London 1657

The ad explains “The Vertue of the COFFEE drink” what coffee is, how it grows, how it cures numerous maladies, including Dropsy, Gout, and Scurvy, …
Long history....

Japan, 1806

USA, 1890
Brand advertising
Goal: create a distinct favorable image
Direct marketing
Advertising that involves a "direct response": buy, subscribe, vote, donate, etc, now or soon
Why “computational” advertising?
Lots of computational this and that …

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

All are about

a) Mixing an old science with **computing capabilities**

b) **Thinking** algorithmically about an old challenge
Why go “computational”? 

- Classical: 
  - Relatively few venues – magazines, billboards, newspapers, handbills, TV, etc 
  - High cost per venue ($3Mil 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
"Half the money I spend on advertising is wasted; the trouble is I don't know which half."

John Wanamaker, ~ 1875
Computational advertising – the central 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

- Related challenge 1: Design markets and exchanges that help in this task, and maximize value for users, advertisers, and publishers
- Related challenge 2: Build the infrastructure to support this process
The central challenge decomposed

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

1. **Representation** = represent the user, the context, and the ads in an **effective & efficient** way
2. **Definition** = define the mathematical optimization problem to capture the actual marketplace constraints and goals
3. **Solution** = solve the optimization problem in an **effective & efficient** way
Computational Advertising and Market Design

MARKET DESIGN

 COMPUTATIONAL ADVERTISING

SYSTEMS
The Opportunity

The money and beyond
Online advertising is a huge business

Online Ad Spending Worldwide, 2010-2015
billions and % change

Note: includes banner ads, classifieds, email (embedded ads only), lead generation, rich media, search, sponsorships and video; includes mobile ads within the existing formats
Source: eMarketer, June 2011

128746
Total worldwide advertising spend

Total Media Ad Spending Worldwide, 2010-2015

Billions and % change

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Media Ad Spending</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>$475.7</td>
<td>5.8%</td>
</tr>
<tr>
<td>2011</td>
<td>$496.9</td>
<td>4.5%</td>
</tr>
<tr>
<td>2012</td>
<td>$529.5</td>
<td>6.6%</td>
</tr>
<tr>
<td>2013</td>
<td>$552.5</td>
<td>4.4%</td>
</tr>
<tr>
<td>2014</td>
<td>$579.8</td>
<td>4.9%</td>
</tr>
<tr>
<td>2015</td>
<td>$603.1</td>
<td>4.0%</td>
</tr>
</tbody>
</table>

Note: includes directories, internet, magazines, newspapers, outdoor, radio and TV
Source: eMarketer, June 2011
US Online Ad Spending

US Online Ad Spending, 2009-2015
billions and % change

Note: includes banner ads (static display), search ads (paid listings, contextual text links & paid inclusion), rich media, video (in-stream, in-banner, in-text), classified ads, sponsorships, lead generation (referrals) & email (embedded ads only); excludes mobile ad spending
Source: eMarketer, March 2011

125987
Hot off the press!

**US Online Ad Revenues, H1 2000-H1 2011**

<table>
<thead>
<tr>
<th>Year</th>
<th>H1</th>
<th>H2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>$4.01</td>
<td>$4.07</td>
</tr>
<tr>
<td>2001</td>
<td>$3.72</td>
<td>$3.41</td>
</tr>
<tr>
<td>2002</td>
<td>$2.98</td>
<td>$3.03</td>
</tr>
<tr>
<td>2003</td>
<td>$3.29</td>
<td>$3.98</td>
</tr>
<tr>
<td>2004</td>
<td>$4.60</td>
<td>$5.03</td>
</tr>
<tr>
<td>2005</td>
<td>$5.79</td>
<td>$6.76</td>
</tr>
<tr>
<td>2006</td>
<td>$7.91</td>
<td>$8.97</td>
</tr>
<tr>
<td>2007</td>
<td>$9.99</td>
<td>$11.21</td>
</tr>
<tr>
<td>2008</td>
<td>$11.51</td>
<td>$11.94</td>
</tr>
<tr>
<td>2009</td>
<td>$10.90</td>
<td>$11.76</td>
</tr>
<tr>
<td>2010</td>
<td>$12.13</td>
<td>$13.91</td>
</tr>
<tr>
<td>2011</td>
<td>H1</td>
<td>$14.94</td>
</tr>
</tbody>
</table>

Source: Interactive Advertising Bureau (IAB) and PricewaterhouseCoopers (PwC), "IAB Internet Advertising Revenue Report: 2011 First Six Months Results," Sep 28, 2011
Total US advertising spend

US Ad Spending & Growth, 2009-2014
(billions and % growth)

- 2009: $163.6
- 2010: $168.5
- 2011: $170.5
- 2012: $178.2
- 2013: $181.8
- 2014: $188.5

Spending vs. % Growth

eMarketer.com
Spending by format

<table>
<thead>
<tr>
<th></th>
<th>H1 2010</th>
<th>H1 2011</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>$5,747</td>
<td>$7,286</td>
<td>26.8%</td>
</tr>
<tr>
<td>Display</td>
<td>$4,356</td>
<td>$5,535</td>
<td>27.1%</td>
</tr>
<tr>
<td>—Banner ads</td>
<td>$2,744</td>
<td>$3,414</td>
<td>24.4%</td>
</tr>
<tr>
<td>—Rich media</td>
<td>$743</td>
<td>$763</td>
<td>2.7%</td>
</tr>
<tr>
<td>—Digital video</td>
<td>$627</td>
<td>$891</td>
<td>42.1%</td>
</tr>
<tr>
<td>—Sponsorship</td>
<td>$242</td>
<td>$467</td>
<td>93.0%</td>
</tr>
<tr>
<td>Classifieds</td>
<td>$1,262</td>
<td>$1,237</td>
<td>-2.0%</td>
</tr>
<tr>
<td>Lead generation</td>
<td>$642</td>
<td>$805</td>
<td>25.4%</td>
</tr>
<tr>
<td>Email</td>
<td>$120</td>
<td>$79</td>
<td>-34.2%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$12,127</strong></td>
<td><strong>$14,942</strong></td>
<td><strong>23.2%</strong></td>
</tr>
</tbody>
</table>

Source: Interactive Advertising Bureau (IAB) and PricewaterhouseCoopers (PwC), "IAB Internet Advertising Revenue Report: 2011 First Six Months Results," Sep 28, 2011

www.eMarketer.com
Who is spending the money?

<table>
<thead>
<tr>
<th>Industry Category</th>
<th>H1 2010</th>
<th>H1 2011</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail</td>
<td>$2,425</td>
<td>$3,437</td>
<td>41.7%</td>
</tr>
<tr>
<td>Telecom</td>
<td>$1,698</td>
<td>$2,092</td>
<td>23.2%</td>
</tr>
<tr>
<td>Financial services</td>
<td>$1,455</td>
<td>$1,942</td>
<td>33.5%</td>
</tr>
<tr>
<td>Auto</td>
<td>$1,334</td>
<td>$1,644</td>
<td>23.2%</td>
</tr>
<tr>
<td>Computing products</td>
<td>$1,213</td>
<td>$1,494</td>
<td>23.2%</td>
</tr>
<tr>
<td>Consumer packaged goods</td>
<td>$970</td>
<td>$897</td>
<td>-7.6%</td>
</tr>
<tr>
<td>Leisure travel</td>
<td>$849</td>
<td>$1,195</td>
<td>40.8%</td>
</tr>
<tr>
<td>Pharma &amp; healthcare</td>
<td>$606</td>
<td>$598</td>
<td>-1.4%</td>
</tr>
<tr>
<td>Media</td>
<td>$485</td>
<td>$598</td>
<td>23.2%</td>
</tr>
<tr>
<td>Entertainment</td>
<td>$485</td>
<td>$598</td>
<td>23.2%</td>
</tr>
<tr>
<td>Other</td>
<td>$606</td>
<td>$448</td>
<td>-26.1%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>$12,127</strong></td>
<td><strong>$14,942</strong></td>
<td><strong>23.2%</strong></td>
</tr>
</tbody>
</table>

Note: industry definitions may have changed over the time period depicted, both within the survey process and definitively by survey respondents; numbers do not add up to 100% because minor categories are not displayed.

Source: Interactive Advertising Bureau (IAB) and PricewaterhouseCoopers (PwC), "IAB Internet Advertising Revenue Report: 2011 First Six Months Results," Sep 28, 2011
Share of Time Spent per Day with Major Media by US Adults, 2008-2010

% of total

2008: 635 mins
2009: 650 mins
2010: 660 mins

- **TV and video:**
  - 40.0% in 2008
  - 41.0% in 2009
  - 40.0% in 2010

- **Internet:**
  - 7.5% in 2008
  - 7.0% in 2009
  - 7.0% in 2010

- **Mobile:**
  - 6.0% in 2008
  - 6.0% in 2009
  - 7.5% in 2010

- **Newspapers:**
  - 21.5% in 2008
  - 22.5% in 2009
  - 23.5% in 2010

- **Magazines:**
  - 16.0% in 2008
  - 15.0% in 2009
  - 14.5% in 2010

Note: time spent with each medium includes all time spent with that medium, regardless of multitasking; for example, 1 hour of multitasking on the internet and watching TV was counted as 1 hour for TV and 1 hour for internet.

Source: eMarketer, Dec 2010

**Internet Today:**
- 23.5% of time
- 15.3% of dollars
- 8.2 p.p. gap
Over Half of Consumers are Simultaneously Online and Watching TV

M-F 6p-9p / Adults 25-54

57% Simultaneous Internet + TV

Media Consumption Level By Usage Quintile

Source: 1 Knowledge Networks SRI, Fall studies, 2 Nielsen TV/Internet Convergence Research Panel
# Dominant players

<table>
<thead>
<tr>
<th></th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>$7.90</td>
<td>$10.03</td>
<td>$12.77</td>
<td>$16.53</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>$3.66</td>
<td>$3.47</td>
<td>$3.46</td>
<td>$3.57</td>
</tr>
<tr>
<td>Facebook</td>
<td>$0.56</td>
<td>$1.21</td>
<td>$2.19</td>
<td>$2.87</td>
</tr>
<tr>
<td>Microsoft</td>
<td>$1.18</td>
<td>$1.49</td>
<td>$1.92</td>
<td>$2.66</td>
</tr>
<tr>
<td>AOL</td>
<td>$0.99</td>
<td>$0.88</td>
<td>$0.85</td>
<td>$0.86</td>
</tr>
<tr>
<td><strong>Total top 5</strong></td>
<td>$14.28</td>
<td>$17.07</td>
<td>$21.19</td>
<td>$26.49</td>
</tr>
<tr>
<td><strong>Total internet</strong></td>
<td>$22.66</td>
<td>$26.04</td>
<td>$31.30</td>
<td>$36.80</td>
</tr>
</tbody>
</table>

*Note: net ad revenues after companies pay traffic acquisition costs (TAC) to partner sites*

*Source: company reports, April-May 2011; eMarketer, Jan & June 2011*
Paid search marketshare

- Worldwide Internet users conduct over 200 billion searches each month
- Google drives roughly 46% of overall online ads with annual growth estimated at +20%
- Search should continue to be driven by leaders Google, Yahoo! and bing
- Search expected to benefit from growth of the mobile Internet

Leading the Search

($ in billions)

Source: Baird Equity Research, Gridley estimates and Search Engine Journal
Why spend on-line?

“Economic unrest means more spending for online advertising—even as ad budgets remain tight. Marketers see online advertising as safe because the ads are typically more **measurable**, more **efficient** and more **targettable**, making the internet a more accountable medium than nearly all traditional channels”

-- E-marketeer, December 29, 2010
The Value of Web Advertising

Advertising supports a vast eco-system on the Web

1. Publisher revenue makes macro & micro publishers viable
2. Focused reach & targeting makes niche interests businesses possible
3. Advertiser revenue makes large scale “free” services possible: Facebook, Google, Twitter, Yahoo, …

- The web would be a lot smaller without advertising ➔ Advertising creates huge direct and indirect value for consumers!
Computational Advertising Landscape
Marketplace 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)
The central challenge decomposed

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

1. **Representation** = represent the user, the context, and the ads in an **effective & efficient** way
2. **Definition** = define the mathematical optimization problem to capture the actual marketplace constraints and goals
3. **Solution** = solve the optimization problem in an **effective & efficient** way
Graphical ads

Yahoo! Sports

Big Blue heaven

PERFECTION belonged to the Giants on Sunday, as New York scored with 35 seconds left to shock New England 17-14 in Super Bowl XLII. Recap Tyree's catch of a lifetime... Peyton pumped... Webcast: Winners & losers...

More on Super Bowl Giants Super Bowl gear Super Bowl history

Bettis, shows no class Our Top 5 Super Bowl ads Online collapses for Patriots

Tom Bellichick couldn’t face the music. MJD blog... Halftime show... Watch the best Super Bowl ads, according to Jamie McMillan. Blog... The Giants made life miserable for Tom Brady. Story... Clinton Crows

ADVERTISMENT

Go to PepsiStuff.com NOW!
Graphical ads – Guaranteed Delivery

- Two types of online graphical advertising
  - Guaranteed delivery (GD)
  - Performance graphical advertising (non-guaranteed delivery, NGD)
- Guaranteed delivery
  - Contract booked based on targeting attributes of an impression: age, income, location,…
  - Each contract has a duration and a desired number of impressions
- Issues in GD
  - Contract pricing
  - Traffic forecasting
  - Impression allocation to the active contracts
  - …
Performance graphical ads

- Graphical ads can also be placed based on performance – CPM/CPC/CPA
- Assume for the moment
  \( \text{Optimization Problem Definition} = \text{Max CTR} \)
- Matching approaches:
  1. **Reactive**: explore the placement of a particular ad on different pages; for each page observe achieved CTR; once the CTRs are learned, given page, pick the ad with highest observed CTR
  2. **Predictive**: generate features for the ad using related ads (same advertiser), landing page, or advertiser metadata – predict performance based on page and ad features
  3. **Hybrid**: (1) and (2) are complementary and can be combined
Representation

1. **Reactive**: explore the placement of a particular ad on different pages; for each page observe achieved CTR; once the CTRs are learned, given page, pick the ad with highest observed CTR
   ➔ Ads represented by achieved CTR/page + weights

1. **Predictive**: generate features for the ad using related ads (same advertiser), landing page, or advertiser metadata – predict performance based on page and ad features
   ➔ Ads (pages) represented by features of ads (resp. pages) + weights

2. **Hybrid**: (1) and (2) are complementary and can be combined
   ➔ Combined representation
The world of display advertising
Display ad technology evolution

Direct Response Orientation

Tech-driven inventory selling
- Ad server infrastructure
- Ad networks

Creating "markets" for inventory
- Ad exchanges / marketplaces

Managing multiple ad channels
- Supply side technologies
- Data management platforms

Intelligently buying into pools of liquidity
- Demand side technologies
- Trading desks

Engagement driven advertising
- Social
- Mobile
- Display

Brand Orientation

Gridley & Company LLC
Textual Ads
Textual ads

1. Ads driven by search keywords = “**sponsored search**”
   (a.k.a. “keyword driven ads”, “paid search”, “adwords”, etc)
   - Advertiser chooses a
     “bid phrase” = query on which to display
   - Can also subscribe to
     “advanced match” = display me on related queries
     - Needed to achieve volume
     - Huge challenge

2. Ads driven by the content of a web page = “**content match**”
   (a.k.a. “context driven ads”, “contextual ads”, “adsense”, etc)

Textual ads are heavily related to Search and IR
Historical view on textual advertising

- Late 1990s Alta Vista tried the Sponsored Search model
  - Rejected by the early search engine users
- Goto.com (acquired later by Overture) develops a search engine for paid ads
  - Users with commercial interest go to this engine
  - At the peak, a billion dollar business
- Google tries the Sponsored Search model again
  - This time a success
- Advertisers cannot get enough volume
  - Content match to provide more impressions
Textual ads anatomy
Search Yahoo for sigir 2010

Also try: sigir 2010 workshops, more...

home [ACM SIGIR 2010]
Deadline for the Elsevier 2010 App Challenge extended to July 9th (read more) ... 33rd Annual ACM SIGIR Conference. 19-23 July 2010, Geneva, Switzerland. Cochairs: Stephane ...
www.sigir2010.org - Cached

cfp [ACM SIGIR 2010]
SIGIR is the major international forum for the presentation of new ... SIGIR 2010 welcomes contributions related to any aspect of IR theory and foundation, techniques, and ...
www.sigir2010.org/dcku.php?id=cfp - Cached

ACM SIGIR Special Interest Group on Information Retrieval ...
ACM SIGIR addresses issues ranging from theory to user demands in the application of ...
Get ready for SIGIR 2010 in Geneva! SIGIR members Ricardo Baeza-Yates ...
www.sigir.org - Cached

ACM SIGIR - Upcoming Events
SAPMIA 2010 - Social, Adaptive and Personalized Multimedia ... SIGIR 2010 (Geneva, Switzerland) - 33rd Annual ACM SIGIR Conference on Research & Development on Information ...
www.sigir.org/events/events-upcoming.html - Cached
Google campaign

Google AdWords

Campaign: Campaign #1
- Enabled
- Budget: $7.00/day
- Targeting: Search
- All devices
- All languages
- All countries and territories

Ad groups
- Settings
- Ads
- Keywords
- Networks

Keywords:
- sigir
- sigir 2010
- sigir2010
- test retrieval conference
- andrej broder
- sigir geneva
- sigir forum
- avgony gabrilovich

Max. CPC | Clicks | Impr. | CTR | Avg. CPC | Cost | Avg. Pos.
---------|-------|-------|-----|----------|------|----------
$0.40 | 20 | 2,449 | 0.82% | $0.16 | $3.19 | 1.2
$0.40 | 3 | 647 | 0.48% | $0.40 | $1.20 | 1
$0.40 | 1 | 25 | 4.00% | $0.16 | $0.16 | 1
$0.40 | 0 | 108 | 0.00% | $0.00 | $0.00 | 1
$0.40 | 0 | 6 | 0.00% | $0.00 | $0.00 | 1
$1.00 | 0 | 108 | 0.00% | $0.00 | $0.00 | 1
$1.00 | 0 | 3 | 0.00% | $0.00 | $0.00 | 1
$1.00 | 0 | 4 | 0.00% | $0.00 | $0.00 | 1
$0.40 | 0 | 12 | 0.00% | $0.00 | $0.00 | 1
Yahoo! Search

![Performance and Account Summary](image)

- **Performance**
  - Impressions: 314
  - Clicks: 4
  - Cost: $0.56

- **Graph**: Clicks

- **Account Summary**

<table>
<thead>
<tr>
<th>Impressions</th>
<th>CTR (%)</th>
<th>Clicks</th>
<th>Avg. CPC ($)</th>
<th>Cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>314</td>
<td>1.27</td>
<td>4</td>
<td>0.14</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Results 1-1 of 1
**Bing**

A screenshot of Microsoft adCenter showing the interface for managing ad groups. The screen displays settings for an ad group named "Ad group #1", with details such as the ad group status (Active) and default search bid (USD: 0.30). Below, there is a table showing performance data for different keywords, including metrics like clicks, impressions, and conversion rates.

*Note: The interface includes options for changing settings and selecting columns for a more detailed view of the data.*
Facebook
Visible and invisible parts

**Title**
- Tutorial at SIGIR 2010
- Information Retrieval Challenges in Computational Advertising
- research.yahoo.com/tutorials/sigir

**Display URL**
- http://research.yahoo.com/tutorials/sigir10_compadv/

**Bid phrase**: sigir 2010
**Bid**: $1
Destination: the landing page

SIGIR 2010 - Geneva, Switzerland
19-23 July, 2010

Tutorial on Information Retrieval Challenges in Computational Advertising
In conjunction with the 33rd ACM SIGIR Conference
19-23 July, 2010 - Geneva, Switzerland

OVERVIEW

Web advertising supports a large swath of the Internet ecosystem. It brings revenue to countless publishers that rent space on their pages for advertising: from small mom-and-pop shops to major Internet companies. It also provides valuable traffic to numerous commercial Web sites and has fueled the development of Web search engines. Today, Web advertising is increasingly impacting the world outside the Internet by shaping the...
Ad Selection
Participants: Publishers, Advertisers, Users, & “Match maker”
Problem definition: Ad selection objective

- Each participant has its own utility
  1. Advertisers want ROI and volume
  2. User wants relevance
  3. Publisher wants revenue per impressions/search
  4. Ad network wants revenue and growth

- Ad selection: optimize for a goal that balances the utilities of the four participants

- Some tradeoffs are linked with the short term and long term business objectives:
  - Allow for easy adjustments based on periodical changes in objectives
Efficiency requirements: Scale and Cost of Serving

- **The Billions:**
  - Billions of individual ads in sponsored search and content match
  - Billions of unique queries/millions of searches per hour
  - Trillions of page impressions (content match and graphical advertising)
  - Billions of users

- **The Milliseconds:**
  - Requests served while the user ‘waits’: no more than 100ms response time

- **The Money:**
  - Serving each requests require some CPU amount
  - Data usually needs to be in memory
  - Per-request cost needs to be lower than the serving cost
  - Low CTR make this a challenging problem
Textual ad selection
In the beginning: The database approach

- **Thinking of SS as a data base problem**
  - SELECT ads
    FROM ad_table
    WHERE bid_phrase = query

- **Implementation**
  - Sponsored search
    - Match the query to the ad bid phrase (some normalization performed)
    - Advertisers cannot bid on all feasible queries (especially in the tail) → Need advanced match
  - Advanced match → translate the query into bid phrases
    - Very difficult to capture context, relevance, etc.
    - Pricing is misleading – bid on original phrase has little to do with value of AM
  - Content match → bid phrases from pages
    - very difficult to capture context, semantics, relevance, etc.
Textual ad schema

Buy appliances on Black Friday

Kitchen appliances

New Year deals on lawn & garden tools

Advertiser

Account 1

Campaign 1

Ad group 1

Creatives

Bid phrases

{ Miele, KitchenAid, Cuisinart, …}

Account 2

Campaign 2

Ad group 2

...

Can be just a single bid phrase, or thousands of bid phrases (which are not necessarily topically coherent)

Brand name appliances
Compare prices and save money
www.appliances-r-us.com
New old concept: 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 textual ad is an information retrieval problem with multiple, possible contradictory utility functions
Finding the “best ad” as an Information Retrieval (IR) problem

- **Representation**: Treat the ads as documents in IR
  
  [Ribeiro-Neto et al. SIGIR 2005] [Broder et al. SIGIR2007] [Broder et al. CIKM2008]

- **Optimization/solution**: Retrieve the ads by evaluating the query over the ad corpus

**Details**

- Analyze the “query” and extract query-features
  
  Query = full context (content, user profile, environment, etc)

- Analyze the documents (= ads) and extract doc-features

- Devise a scoring function = predicates on q-features and d-features + weights

- Build a search engine that produces quickly the ads that maximize the scoring function
Setting the ad retrieval problem

- **Ads corpus =**
  - Textual Ads: Bid phrase(s) + Title + Creative + URL + Landing Page + …
  - Graphical Ads: Advertiser supplied meta data + URL + Landing Page + Content of pages with clicks…

- **Query features =**
  - Search Keywords + Outside Knowledge Expansion + Context features

- **Context features (for sponsored search) =**
  - Location + User data + Previous searches + …

- **Context features (for content match) =**
  - Location + User data + Page topic + Page keywords …
Summary
Key messages

1. **Computational advertising = A principled way to find the "best match" between a user in a context and a suitable ad.**

2. **Key sub-problems:**
   I. Representation of user/context/ads
   II. Definition of optimization problem
   III. Efficient and effective solution

3. **The financial scale for computational advertising is huge**
   - Small constants matter
   - Expect plenty of further research

4. **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 often a type of information retrieval problem with multiple, possibly contradictory utility functions

5. **New application domains and new techniques are emerging every day**
   - Good area for research + new businesses!
Conferences where most of the action happens

- WWW (World Wide Web)
- WSDM (Web Search and Data Mining)
- SIGIR (Information Retrieval)
- CIKM (Information and Knowledge Management)
- EC (Electronic Commerce)
- 4 workshops in 2009: SIGIR, KDD, EC, WINE
- First to sixth Workshop on Sponsored Search Auctions
Questions?

We welcome suggestions about all aspects of the course: msande239-aut0910-staff
Thank you!

broder@yahoo-inc.com
vanjaj@yahoo-inc.com

http://research.yahoo.com