

2018 EE448, Big Data Mining, Lecture 11

Computational Advertising

Weinan Zhang

Shanghai Jiao Tong University

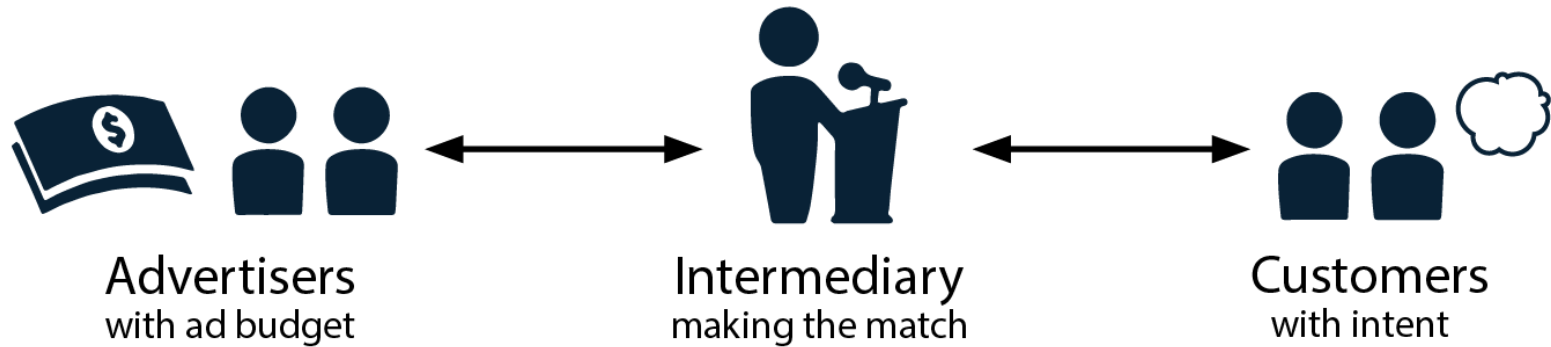
<http://wnzhang.net>

<http://wnzhang.net/teaching/ee448/index.html>

Content of This Course

- Introduction to computational advertising
- Auction for ad selection
- Sponsored search
- Contextual advertising

Advertising



- Make the best match between **advertisers** and **customers** with **economic constraints**



*“Half the money I
spend on advertising is
wasted; the trouble is I
don’t know which
half.”*

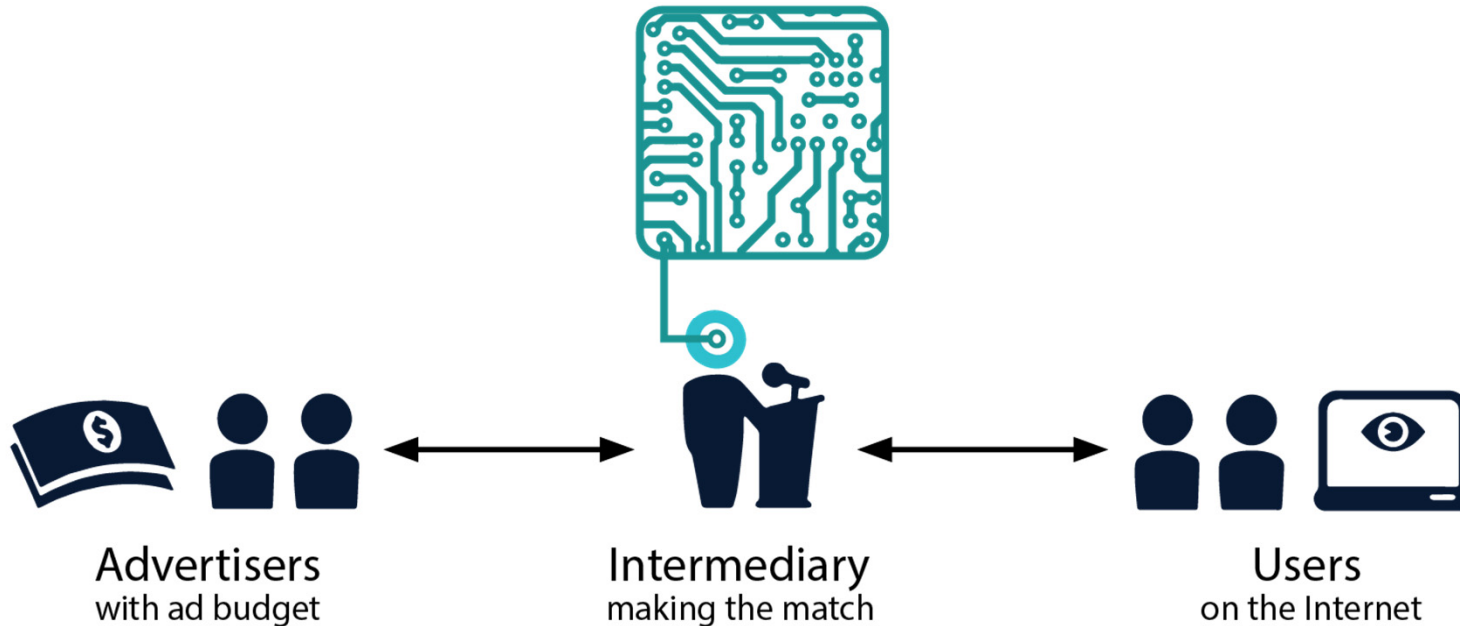
*- John Wanamaker
(1838-1922)*

*Father of modern advertising
and a pioneer in marketing*

Wasteful Traditional Advertising

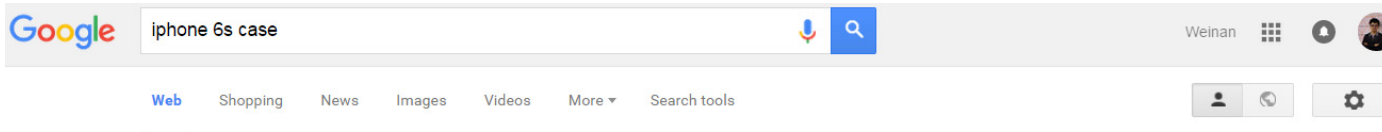


Computational Advertising



- Design **algorithms** to make the best match between the advertisers and Internet users with economic constraints

Sponsored Search



About 16,900,000 results (0.33 seconds)

iPhone 6s Cases - case-mate.com
Ad www.case-mate.com/iPhone-6s-Cases
4.6 ★★★★★ rating for case-mate.com
Shop The iPhone 6s Case Collection. Free Standard Shipping!
Refined Protection · Slim & Tough · Genuinely Crafted · Premium Designs

iPhone 6s
Ad www.apple.com/
The only thing that's changed is everything. Learn more.
A9 chip · Two sizes · Now in rose gold
Pre-order 9.12 - iPhone Upgrade Program - 3D Touch - Cameras

In the news



Speck's iPhone 6s CandyShell + MightyShell cases bring best-of-breed protection to Apple's latest iPhones

9 to 5 Mac - 1 day ago

With the iPhone 6s and iPhone 6s Plus debuting next week, it's important to start thinking ...

Moshi's iPhone 6s and 6s Plus cases offer premium protection
iMore - 23 hours ago

Top 5 Best Leather iPhone 6s Cases
Heavy.com - 12 hours ago

[More news for iphone 6s case](#)

iPhone 6s Cases & Covers from OtterBox

www.otterbox.com/en-us/iphone-6s-cases OtterBox
Get protection that inspires confidence with iPhone 6s cases and covers from OtterBox.
Demandware SiteGenesis.

iPhone 6s - Accessories - Apple

www.apple.com > iPhone > iPhone 6s > Apple Inc.
The essential Apple-designed cases, accessories and all-new aluminum docks for iPhone 6s and iPhone 6s Plus.

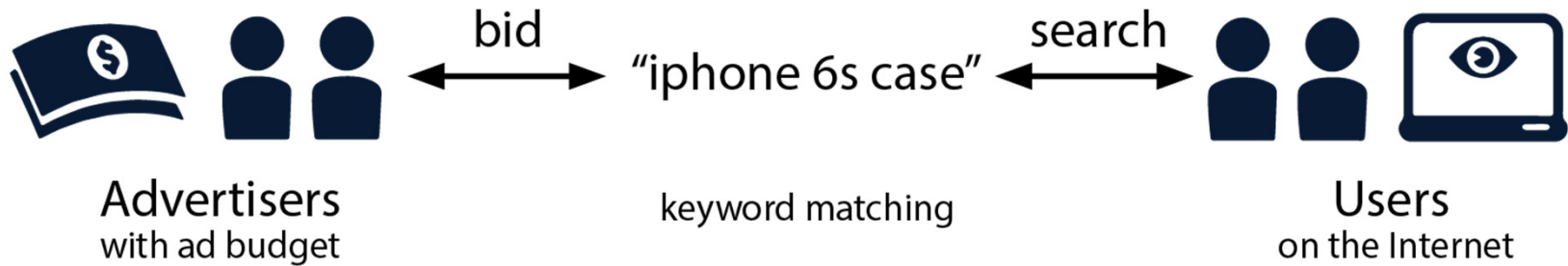
Shop for iphone 6s case on Google

Sponsored ⓘ

 Case-mate - Karat Case Fo... \$49.99 Best Buy ★★★★★ (163)	 Moshi - Iglaze Armour Case... \$39.99 Best Buy ★★★★★ (161)	 Logitech - Protection... \$21.99 Best Buy ★★★★★ (90)	 Moshi - Overture Wall... \$49.99 Best Buy ★★★★★ (18)
 Case-mate - Brilliance Cas... \$44.99 Best Buy ★★★★☆ (294)	 Case-mate - Wallet Folio C... \$54.99 Best Buy ★★★★★ (173)	 Marc by Marc Jacobs Metalli... \$38.00 shopbop	 Case-mate - Karat Hard Sh... \$49.99 Best Buy ★★★★★ (34)

Search: iphone 6s case

Sponsored Search



- Advertiser sets a bid price for the keyword
- User searches the keyword
 - This explicitly shows her information need
- Search engine hosts the auction to ranking the ads

Contextual Advertising

Baidu Workers Call Off Strike

Article

Comments (1)

 SUBSCRIBER CONTENT PREVIEW

FOR FULL SITE ACCESS: [SUBSCRIBE NOW](#)

BEIJING -- Workers at Baidu Inc. in two southern Chinese cities said they called off their strike Monday, although they said hadn't reached any immediate agreement with the Internet search company over complaints about compensation.

Representatives of several hundred employees at the two southern offices of Baidu, which runs China's top search site, finished meetings on Monday with executives from the company's Beijing headquarters, including the head of human resources and an assistant to Robin Li, the company's chief executive. They plan to give the company two to three days to come up with a fair response, and will decide ...

TO CONTINUE READING, [SUBSCRIBE NOW](#) »

 Email  Printer Friendly  Order Reprints

Article Tools
Sponsored by 

Share:        

Sponsored Links

How I Cured My Wrinkles

See How I Got Rid Of Wrinkles With My 2 Step System. Shocking Report!

HowToFixYourWrinkles.com

Bad Credit Guide

Looking to find bad credit? See our comprehensive guide.

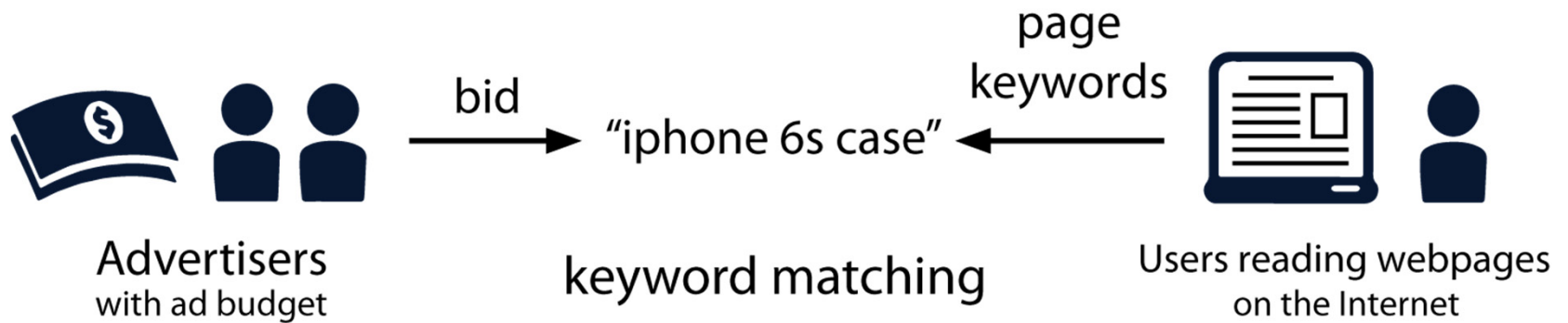
Tradewestsales.Com

▶ I Cured My Wrinkles ◀

I Used (2) Secret Products To Get Rid Of My Wrinkles Forever.

www.GracesWrinkles.com

Contextual Advertising



- Advertiser sets a bid price for the keyword
- Search engine extracts topic keywords from webpages
 - Assuming the user's information need is the webpage content
- Search engine hosts the auction to ranking the ads

Display Advertising

Exxon Mobil Investigated in New York Over Possible Lies on Climate

By JUSTIN GILLIS and CLIFFORD KRAUSS
3:30 PM ET

The sweeping inquiry, by the state attorney general, focuses on whether the oil company lied to the public and investors over the risks of climate change.

250 Comments



T. Fallon/Bloomberg, via Getty Images

An Exxon Mobil refinery in Los Angeles, Calif. The New York attorney general is investigating the oil and gas company.

LATEST NEWS

- 5:01 PM ET 'Grand Theft Auto' Maker Take-Two's Revenue Nearly Triples
- 5:00 PM ET United Airlines CEO to Return in Early 2016 After Heart Attack
- 4:57 PM ET NY Attorney General Investigating Exxon Over Climate Statements

MARKETS »

At close 11/05/2015

European Union Predicts Economic Gains From Influx of Migrants

By JAMES KANTER
12:10 PM ET

Officials forecast that the three million arrivals expected by 2017 would provide a net gain of perhaps a quarter of 1 percent by that year to the European economy.



INSIGHT & ANALYSIS

COMMON SENSE

Dewey Jury's Deadlock Exposes a System's Flaws

By JAMES B. STEWART
3:06 PM ET

One reason for the mistrial in the Dewey & LeBoeuf criminal case may have been the requirement for a unanimous decision.

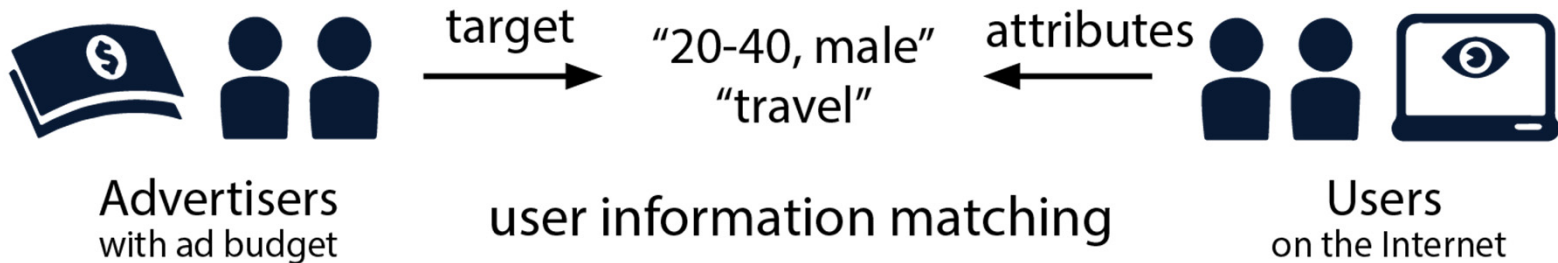


 **BACKBASE**

Backbase a Leader in the Forrester Wave for Omni-Channel Digital Banking

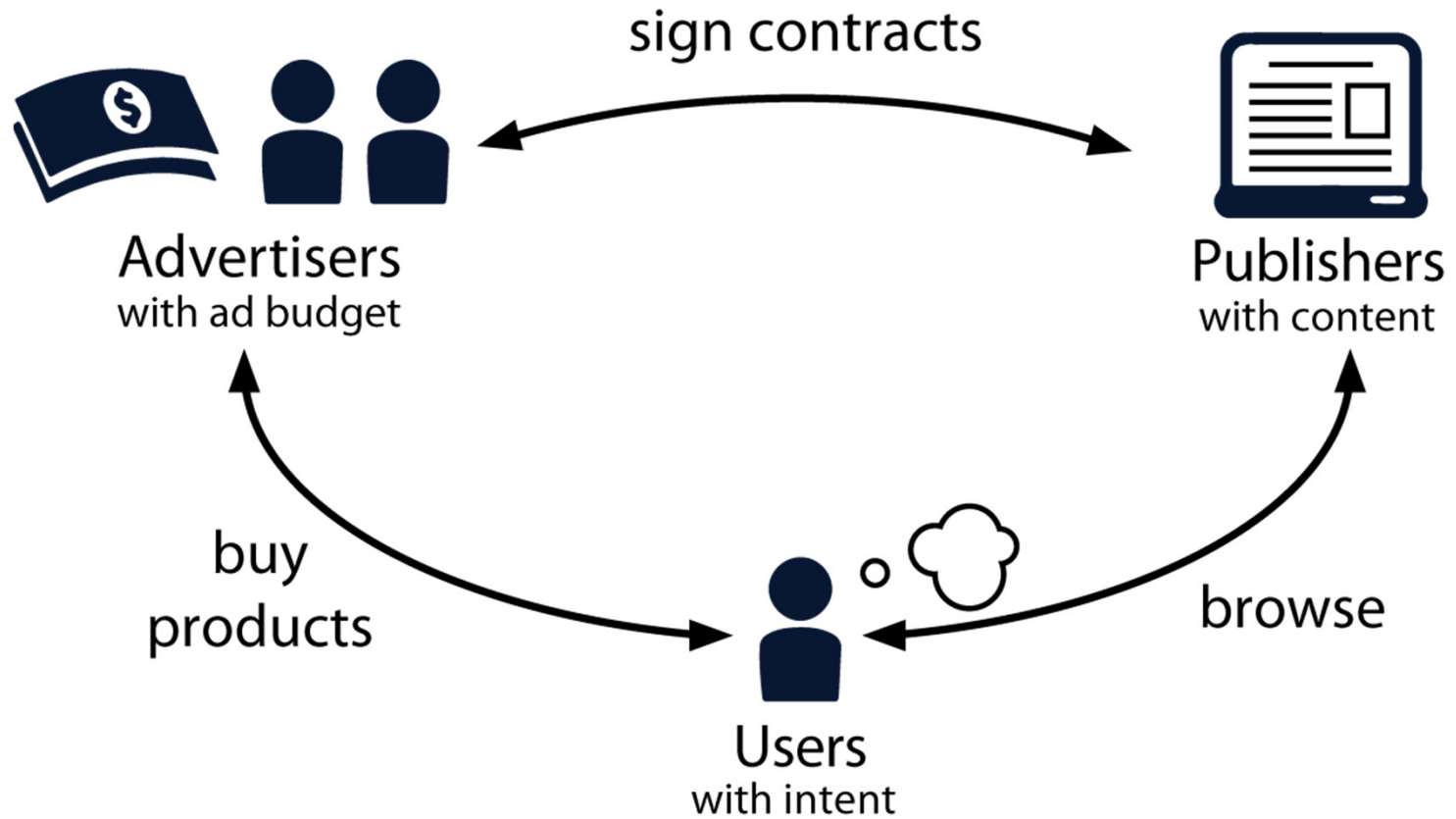
Read the Report

Display Advertising



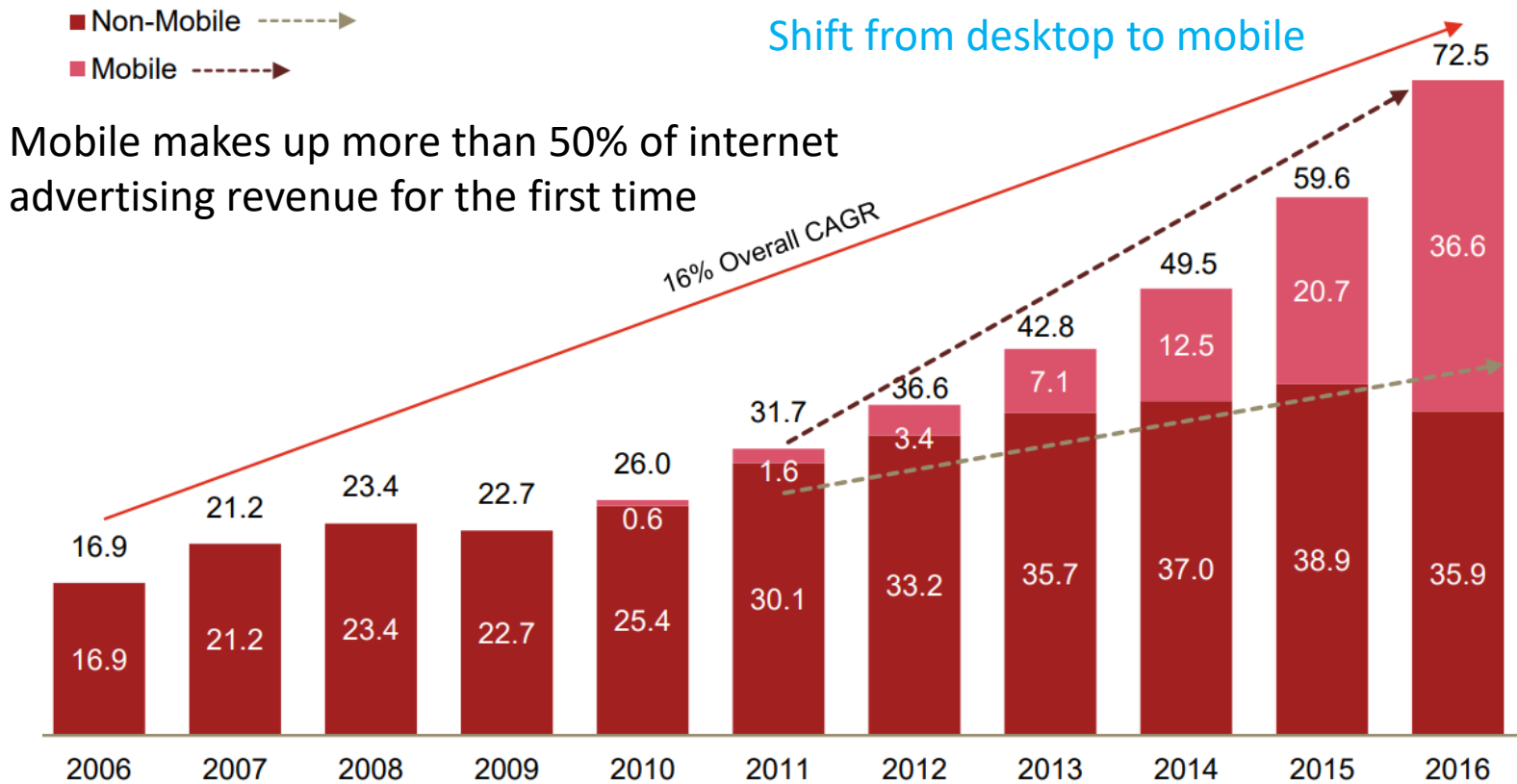
- Advertiser targets a segment of users
 - No matter what the user is searching or reading
- Intermediary matches users and ads by user information

A 3-Player Game



Computational Advertising Markets

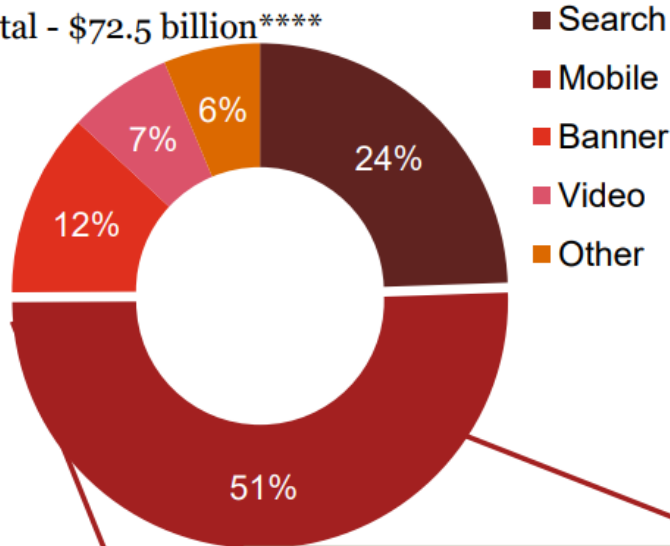
- Statistics from IAB 2016 annual report



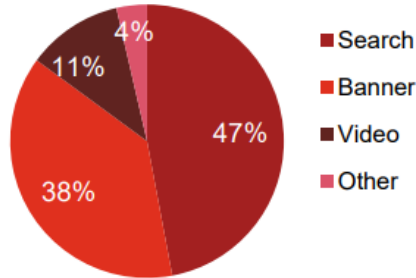
Computational Advertising Markets

Ad formats – full year 2016

Total - \$72.5 billion****

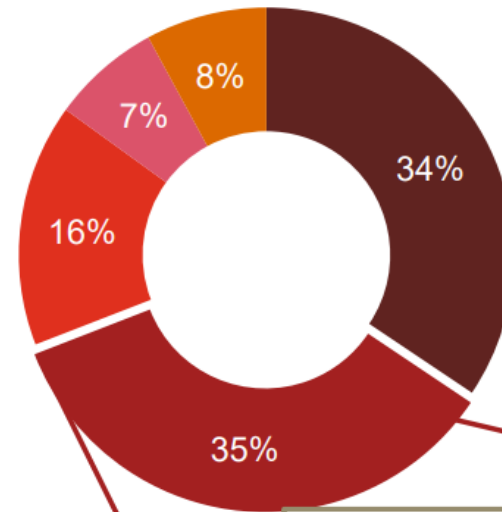


Mobile Formats

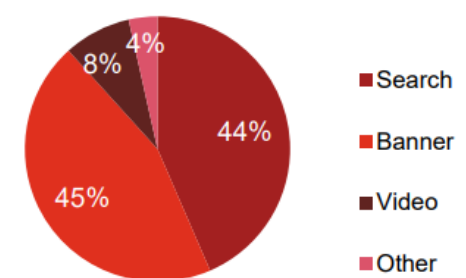


Ad formats – full year 2015

Total - \$59.6 billion*****



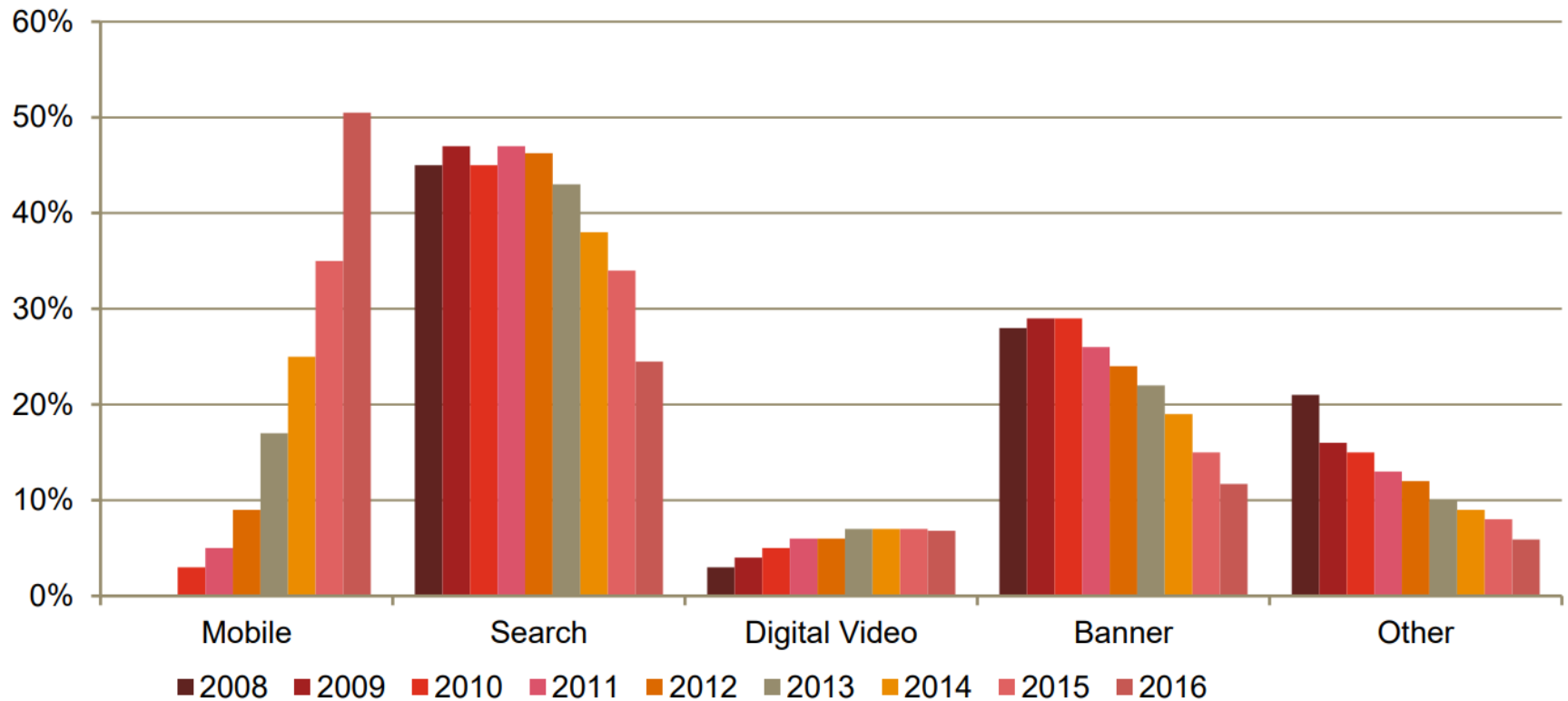
Mobile Formats



Source: IAB/PwC Internet Ad Revenue Report, FY 2016

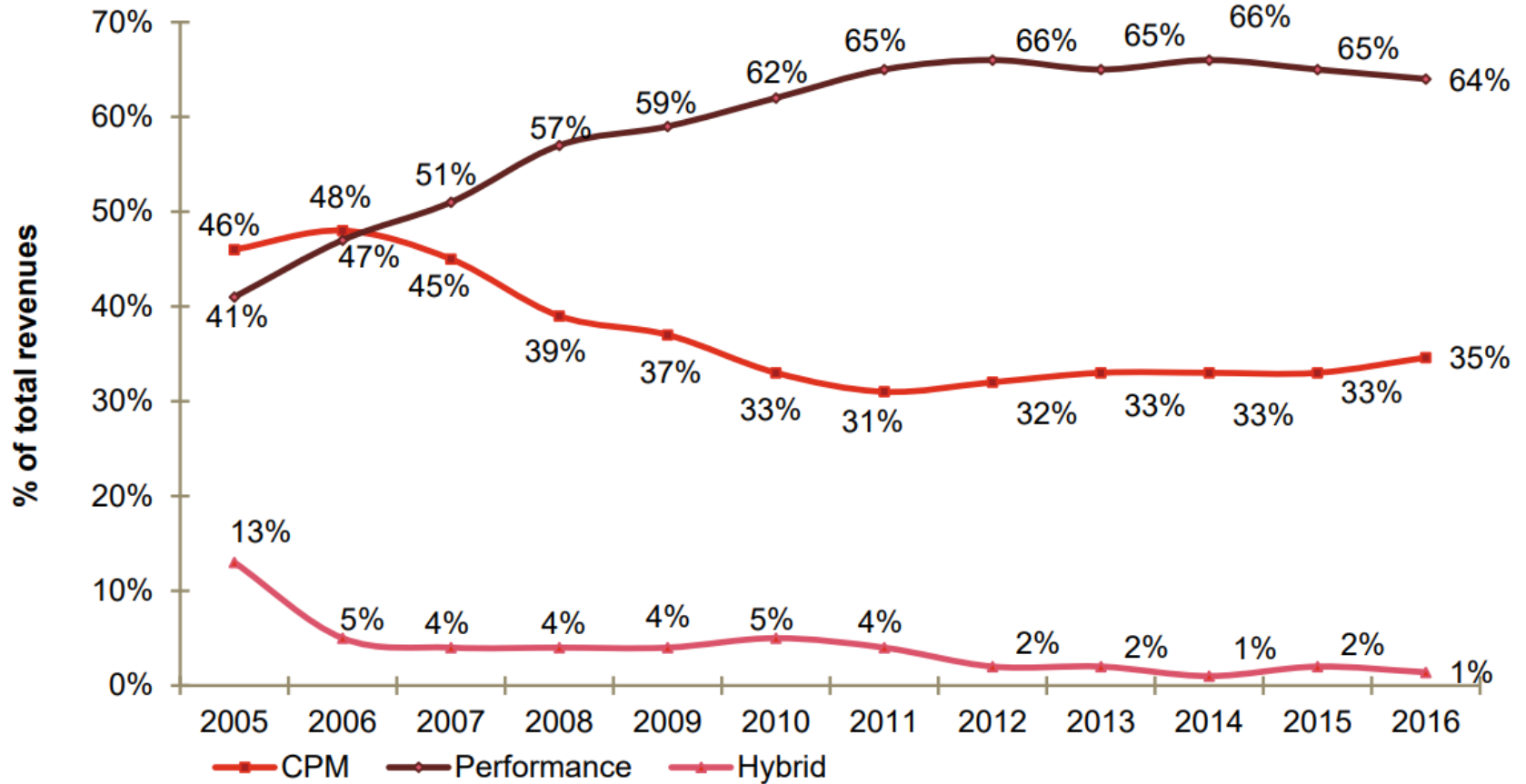
Computational Advertising Markets

Advertising format share, 2008 - 2016* (% of total revenue)



Computational Advertising Markets

Internet ad revenues by pricing model*



Advertisers care more and more about the ad performance, which drives a high motivation of data science for computational advertising optimization.

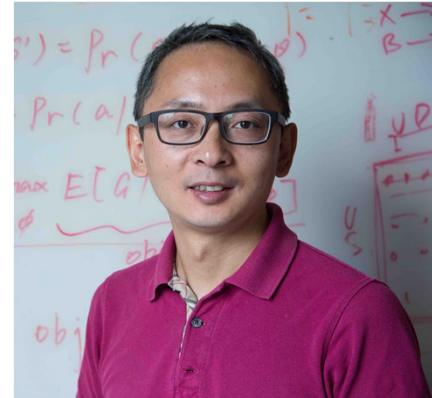
Computational Advertising Experts



Dr. Andrei Broder
Google



Dr. Vanja Josifovski
Pinterest



Prof. Jun Wang
University College London

- Stanford MS&E 239: Introduction to Computational Advertising
 - By Andrei Broder and Vanja Josifovski
 - <https://web.stanford.edu/class/msande239/>
- UCL COMPM041: Web Economics
 - By Emine Yilmaz and Jun Wang
 - http://www.cs.ucl.ac.uk/current_students/syllabus/comp gw/comp gw02_web_economics/
- Some material of this lecture is borrowed from these masters

Content of This Course

- Introduction to computational advertising
- Auction for ad selection
- Sponsored search
- Contextual advertising

Online Auctions

- An auction is a process of buying and selling goods or services by offering them up for bid, taking bids, and then selling the item to the highest bidder
- Auctions are popular
 - Historical sale tool
 - Bonds, treasury bills, land leases, privatization, art, etc.
 - Internet marketplace
 - eBay changed the landscape as a gigantic auctioneer
 - Sponsored search (Google, Facebook, etc.)
 - Display ad exchange (Google AdX, Taobao's TANX, etc.)

Auction Settings



- Imagine we want to sell a **single** item
 - Later we'll extend this to multiple items
- We don't know what it's generally worth
 - Just what it's worth to us
 - We may set a **reserve price** (later)
- Each **bidder** (player) has her own intrinsic value of the item.
 - Willing to purchase it up to this price
 - Values are independent
- But we don't know these values
 - Differs from our previous game theory assumptions about knowledge of payoffs
- How should we proceed?

First Steps

- We could just ask how much people are willing to pay
 - But would they lie?
 - Or manipulate the outcome?
- Problem:
 - How do we motivate buyers to reveal their true values?
- Auction theory: a sub-field of Mechanism Design
 - We design the market, “Economists as engineers”
 - Design an auction so that in equilibrium we get the results that we want

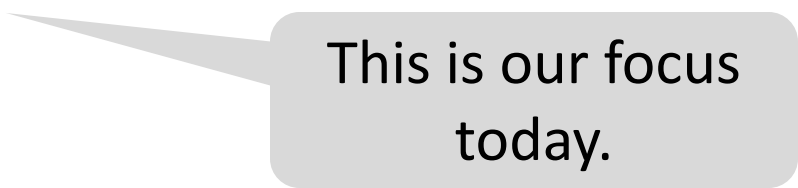


Goal of Auctions

A seller (“auctioneer”) may have several goals.

Most common goals:

1. Maximize **revenue** (profit)
2. Maximize **social welfare** (efficiency)
 - Give the item to the buyer that wants it the most.
(regardless of payments.)
3. **Fairness**
for example, give items to the poor.



This is our focus today.

First Price Auction

- Each bidder writes his bid in a sealed envelope.
- The seller:
 - Collects bids
 - Open envelopes.
- Winner: the bidder with the highest bid.
- Payment: the winner pays her bid.
- Note: bidders do not see the bids of the other bidders.



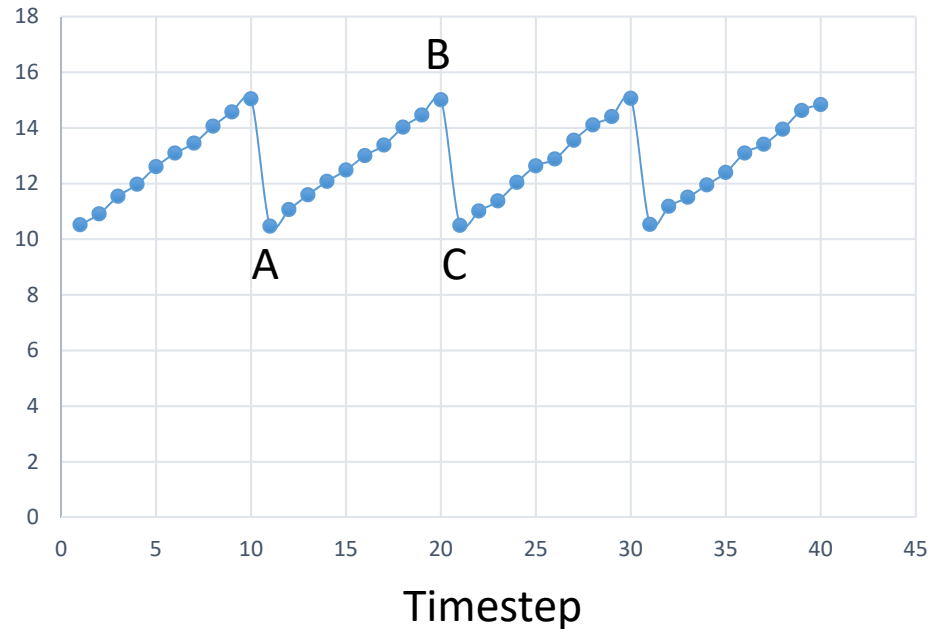
at \$8



First Price Auction is Unstable

- The constant price experimentation led to prices for all queries being updated essentially all the time (why?)
- This resulted in a highly turbulent market

Highest Price



Lead to buyer's remorse and gaming

- A -> B: two bidders raise bid prices to get the first position
- B -> C: One of them reaches its maximum and then goes for the second position

Second Price Auction

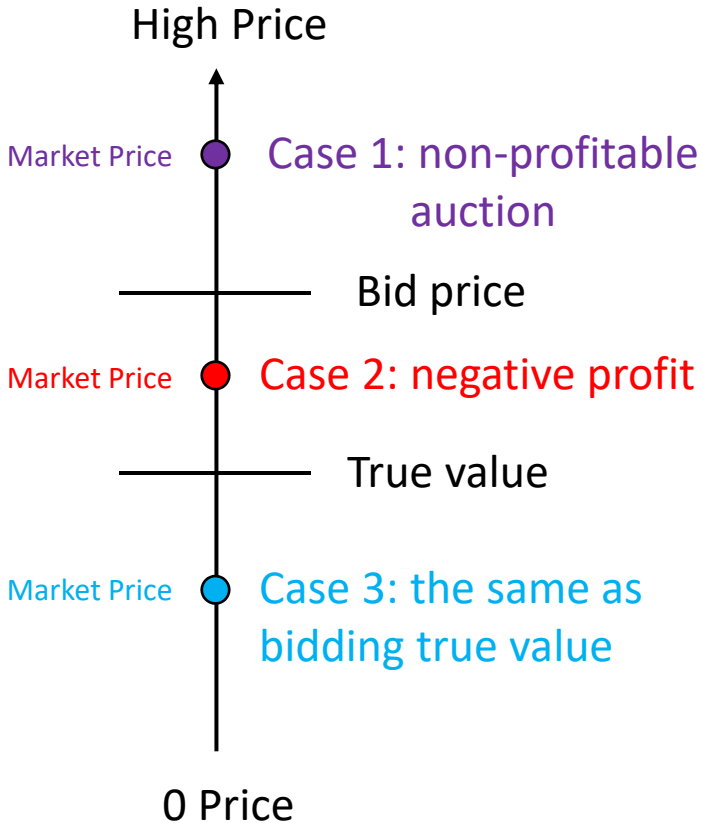
- Each bidder writes his bid in a sealed envelope.
- The seller:
 - Collects bids
 - Open envelopes.
- Winner: the bidder with the highest bid.
- Payment: the winner pays the **second** highest bid.
- Note: bidders do not see the bids of the other bidders.



at \$5



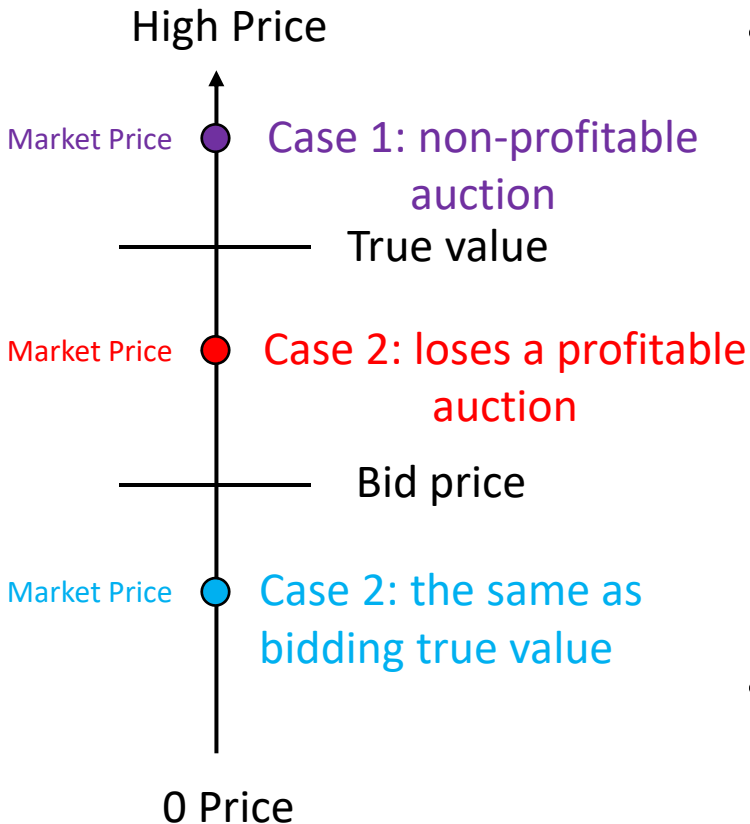
Second Price Auction is Truth-Telling



- If a bidder bids higher than her true value, then
 1. $\text{Market price} > \text{true value} > \text{bid}$: Not a profitable auction
 2. $\text{Bid} > \text{market price} > \text{true value}$: win a negative profit
 3. $\text{Bid} > \text{true value} > \text{market price}$: the same as bidding true value
- Therefore, bidding higher than the true value is less optimal than bidding the true value

Market Price: the highest bid from all other competitors

Second Price Auction is Truth-Telling



- If a bidder bids lower than her true value, then
 1. Market price > true value > bid:
Not a profitable auction
 2. Market price > bid > true value:
she loses a profitable auction
 3. True value > bid > market price:
the same as bidding true value
- Therefore, bidding lower than the true value is less optimal than bidding the true value

Market Price: the highest bid from all other competitors

A Mathematic Proof

- Notations

- Market price as a random variable z
- True value r

Profit given a bid b :
$$R(b) = \int_0^b (r - z)p(z)dz$$

Optimal bid:
$$b^* = \max_b R(b)$$

$$\frac{\partial R(b)}{\partial b} = (r - b)p(b)$$

$$\frac{\partial R(b)}{\partial b} = 0 \Rightarrow b^* = r$$

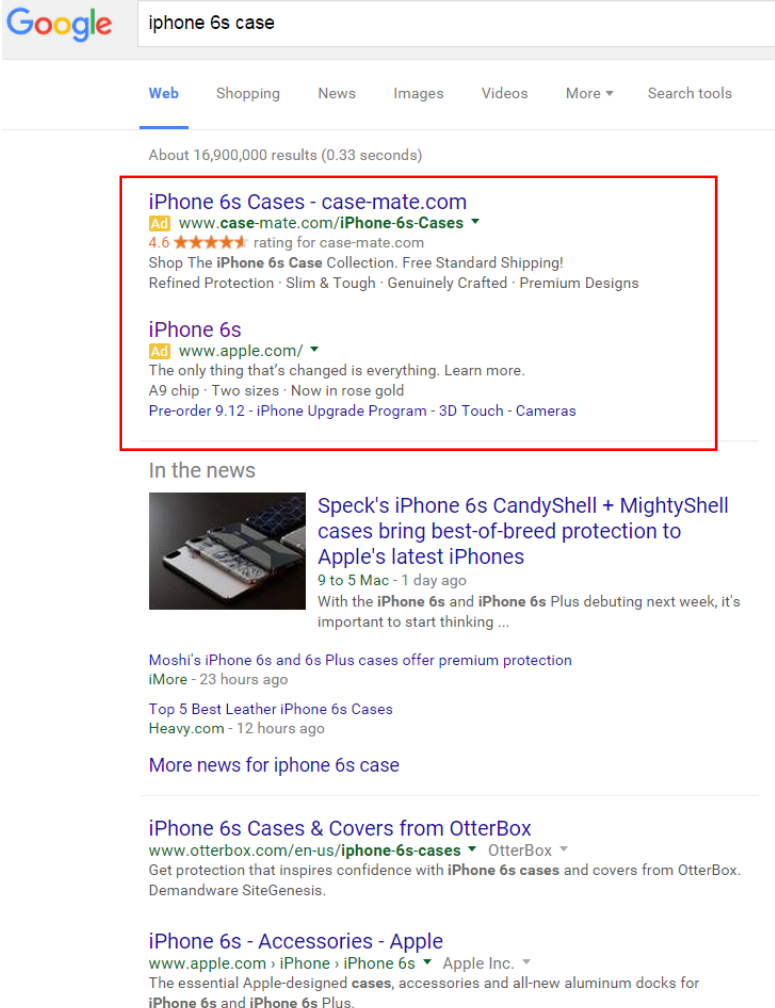
i.e., bid true value

Content of This Course

- Introduction to computational advertising
- Auction for ad selection
- Sponsored search
- Contextual advertising

Sponsored Search

- Sponsored search, when a consumer searches for a term using a search engine, the advertisers' webpage appears as **sponsored links** next to the **organic search results** that would otherwise be returned using the neutral criteria employed by the search engine.
- Pricing scheme: cost-per-click (CPC), i.e., the advertiser pays the search engine a cost for each of the users' clicks on the ad link



The screenshot shows a Google search for "iphone 6s case". The search bar is at the top with the Google logo. Below the search bar are navigation tabs for Web, Shopping, News, Images, Videos, More, and Search tools. The search results show "About 16,900,000 results (0.33 seconds)".

The first result is a sponsored link for "iPhone 6s Cases - case-mate.com". It includes a red box around the ad text, which contains the following information:

- Ad** www.case-mate.com/iPhone-6s-Cases
- 4.6 ★★★★★ rating for case-mate.com
- Shop The iPhone 6s Case Collection. Free Standard Shipping!
- Refined Protection · Slim & Tough · Genuinely Crafted · Premium Designs

Below the sponsored link is another sponsored link for "iPhone 6s" from Apple, with the following information:

- Ad** www.apple.com/
- The only thing that's changed is everything. Learn more.
- A9 chip · Two sizes · Now in rose gold
- Pre-order 9.12 · iPhone Upgrade Program · 3D Touch · Cameras

Below the sponsored links is a section titled "In the news". It features a small image of an iPhone 6s case and a news article titled "Speck's iPhone 6s CandyShell + MightyShell cases bring best-of-breed protection to Apple's latest iPhones". The article is dated "9 to 5 Mac - 1 day ago" and includes the text "With the iPhone 6s and iPhone 6s Plus debuting next week, it's important to start thinking ...".

Below the news section are several more sponsored links:

- Moshi's iPhone 6s and 6s Plus cases offer premium protection (iMore - 23 hours ago)
- Top 5 Best Leather iPhone 6s Cases (Heavy.com - 12 hours ago)
- More news for iphone 6s case
- iPhone 6s Cases & Covers from OtterBox (www.ottterbox.com/en-us/iphone-6s-cases)
- iPhone 6s - Accessories - Apple (www.apple.com)

Players in Sponsored Search

- Advertisers
 - Submit ads associated to certain bid keywords
 - Bid for positions
 - Pay CPC
- Users
 - Make queries to search engine, expressing some intent
- Search engine (a special case of publishers)
 - Executes query against web corpus and other data sources
 - Executes query against the ad corpus
 - Displays a search result page, i.e., organic results + ads

Player Utilities

- Each of the SE, Advertisers, and Users has its own utility
 - Advertisers: maximize the profit from advertising
 - Profit = product revenue – ad cost
 - Users: efficiently find the information they need
 - No matter the organic webpages or ad links
 - Search engine: maximize the profit by displaying ads
 - Profit = Probability of user click × CPC
- Whether is there are win-win-win solution?

Key Questions in Sponsored Search

- Key question 1: how to quantitatively estimate whether the user would like the displayed ad?
 - Solution: user click-through rate (CTR) estimation by machine learning
- Key question 2: how to rank and charge the ads given the bids and estimated CTRs?
 - Solution: generalized second price (GSP) auction

REVIEW

User Response Estimation Problem

- Problem definition

One instance data

- Date: 20160320
- Hour: 14
- Weekday: 7
- IP: 119.163.222.*
- Region: England
- City: London
- Country: UK
- Search Query: "iphone 6s case"
- OS: Windows
- Browser: Chrome
- Ad title: "iphone 6s case on sale!"
- Ad content: "Customize your case design"
- Bid keywords: "iphone case"
- User occupation: Student
- User tags: Sports



Corresponding label

Click (1) or not (0)?

Predicted CTR (0.15)

REVIEW

One-Hot Binary Encoding

- A standard feature engineering paradigm

$x = [\text{Weekday=Friday}, \text{Gender=Male}, \text{City=Shanghai}]$

$x = [0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, \dots, 0]$

Sparse representation: $x = [5:1, 9:1, 12:1]$

- High dimensional sparse binary feature vector
 - Usually higher than 1M dimensions, even 1B dimensions
 - Extremely sparse
- More continuous features may be added
 - IR features, e.g., similarity between query and bid keywords, relevance between query and ad landing page

REVIEW

Training/Validation/Test Data

- Examples (in LibSVM format)

```
1 5:1 9:1 12:1 45:1 154:1 509:1 4089:1 45314:1 988576:1
0 2:1 7:1 18:1 34:1 176:1 510:1 3879:1 71310:1 818034:1
...
```

- Training/Validation/Test data split
 - Sort data by time
 - Train:validation:test = 8:1:1
 - Shuffle training data

REVIEW

Training Logistic Regression

- Logistic regression is a binary classification model

$$p_{\theta}(y = 1|x) = \sigma(\theta^{\top} x) = \frac{1}{1 + e^{-\theta^{\top} x}}$$

- Cross entropy loss function with L2 regularization

$$\mathcal{L}(y, x, p_{\theta}) = -y \log \sigma(\theta^{\top} x) - (1 - y) \log(1 - \sigma(\theta^{\top} x)) + \frac{\lambda}{2} \|\theta\|_2^2$$

- Parameter learning

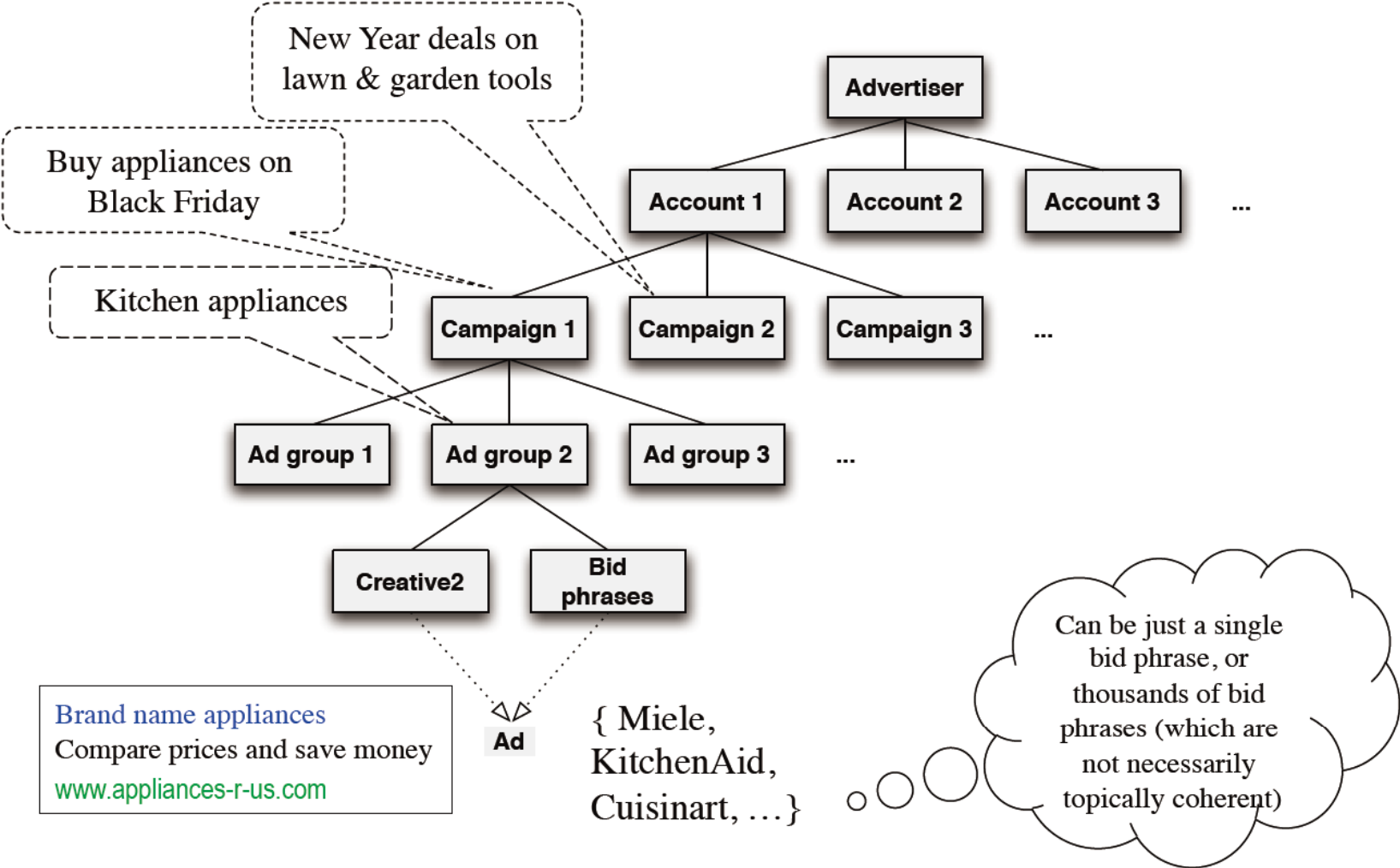
$$\theta \leftarrow (1 - \lambda\eta)\theta + \eta(y - \sigma(\theta^{\top} x))x$$

- Only update non-zero entries

Key Questions in Sponsored Search

- Key question 1: how to quantitatively estimate whether the user would like the displayed ad?
 - Solution: user click-through rate (CTR) estimation by machine learning
- Key question 2: how to rank and charge the ads given the bids and estimated CTRs?
 - Solution: generalized second price (GSP) auction

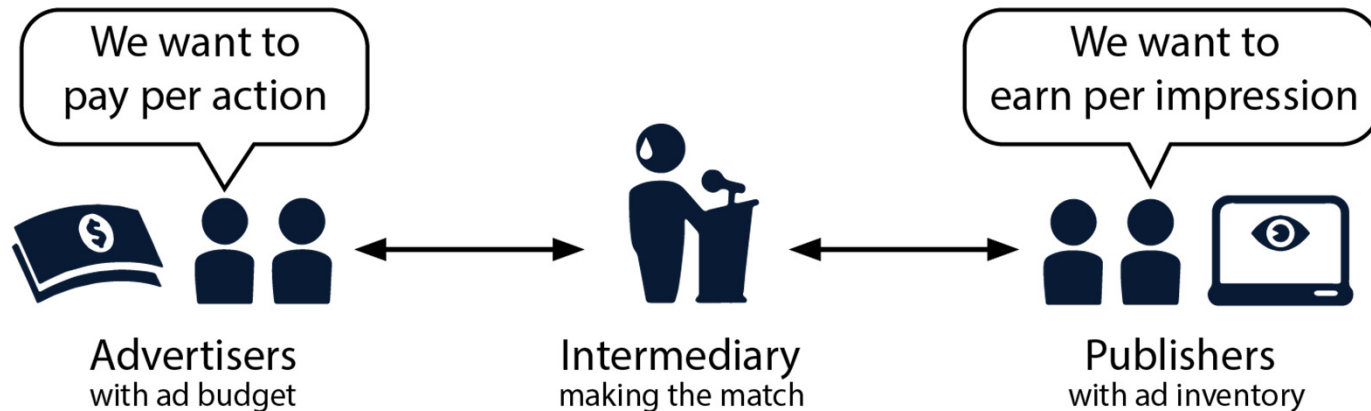
Advertiser Hierarchy



Advertiser-side Keyword Optimization

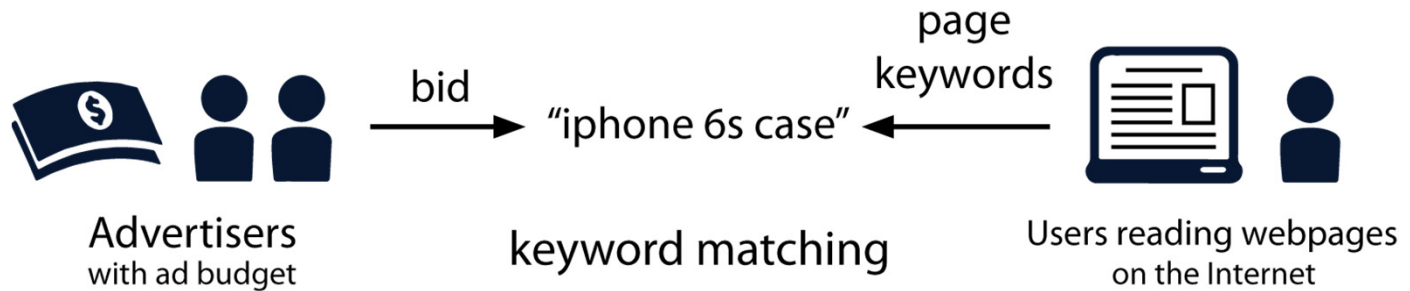
- Recommendation of new keywords to bid
 - Relevance r : the relevance between the keyword and the ad content. Can be measured using IR techniques
 - Volume v : how many search involves such a keyword daily? Can be obtained from search engine logs
 - Competitiveness c : how many advertisers are bidding for this keyword and how high is the market price?
- A straightforward solution: rank the candidate key words by $r \times v / c$
 - i.e., high relevance to the ad, high search volume, low competitiveness

Sponsored Search Ad Pricing Scheme



- Different pricing schemes
 - CPM: cost per mille impression [favored by publisher]
 - charge the advertiser for each ad display
 - CPA: cost per action [favored by advertiser]
 - Charge the advertiser for each conversion (i.e., user execute a particular action, e.g., product purchase, registration etc.)
 - CPC: cost per click [practically used for sponsored search]
 - A trade-off between CPM and CPA

Sponsored Search Ad Ranking



- A straight forward ad ranking scheme: rank the ad by expected cost, i.e., predicted CTR \times bid
- An example candidate ad list

Ad ID	pCTR (%)	Bid (\$)	Expected Cost (\$)	Rank
1	0.3	0.1	0.0003	4
2	0.2	0.8	0.0016	1
3	0.1	0.5	0.0005	3
4	0.3	0.05	0.00015	5
5	0.2	0.4	0.0008	2

Generalized Second Price (GSP) Auction

- Classic second price auction (ranking by bid): the winner pays the second highest bid price
 - i.e., pay the lowest price to maintain her winning situation
- Generalized second price auction
 - ranking by $\text{CTR} \times \text{bid}$
 - There are multiple ad slots
 - the winner of each ad slot pays the lowest price to maintain her winning situation

$$\text{cost}_i = \frac{\text{CTR}_{i+1} \times \text{bid}_{i+1}}{\text{CTR}_i}$$

- GSP is also known as Vickrey-Clarke-Groves (VCG) auctions

An Example of GSP

Ad ID	Rank	pCTR (%)	Bid (\$)	Expected Cost (\$)	Cost (\$)
2	1	0.2	0.8	0.0016	0.4
5	2	0.2	0.4	0.0008	0.25
3	3	0.1	0.5	0.0005	0.3
1	4	0.3	0.1	0.0003	0.05
4	5	0.3	0.05	0.00015	entry price

- To rank higher: the advertiser needs to
 - (1) design ad content, seek bid keywords to higher the CTR
 - (2) higher the bid price
- To pay less: the advertiser needs to make efforts to higher the ad CTR

$$\text{cost}_i = \frac{\text{CTR}_{i+1} \times \text{bid}_{i+1}}{\text{CTR}_i}$$

Content of This Course

- Introduction to computational advertising
- Auction for ad selection
- Sponsored search
- Contextual advertising

Contextual Advertising

- Textual advertising on third party web pages
- Complement the content of the web page with paid content
- Ubiquitous on the web
- Supports the diversity of the web
 - Sites small and big rely on CM revenue to cover for the cost of existence
- Players
 - Google AdSense
 - Microsoft ContentAds
 - Baidu Display Network

Baidu Workers Call Off Strike

Article | Comments (1)

SUBSCRIBER CONTENT PREVIEW


FOR FULL SITE ACCESS: [SUBSCRIBE NOW](#)


BEIJING -- Workers at Baidu Inc. in two southern Chinese cities said they called off their strike Monday, although they said hadn't reached any immediate agreement with the Internet search company over complaints about compensation.

Representatives of several hundred employees at the two southern offices of Baidu, which runs China's top search site, finished meetings on Monday with executives from the company's Beijing headquarters, including the head of human resources and an assistant to Robin Li, the company's chief executive. They plan to give the company two to three days to come up with a fair response, and will decide ...

TO CONTINUE READING, [SUBSCRIBE NOW](#) »

Email | Printer Friendly | Order Reprints

Article Tools Sponsored by 

Share: 

Sponsored Links

How I Cured My Wrinkles
See How I Got Rid Of Wrinkles With My 2 Step System. Shocking Report!
HowToFixYourWrinkles.com

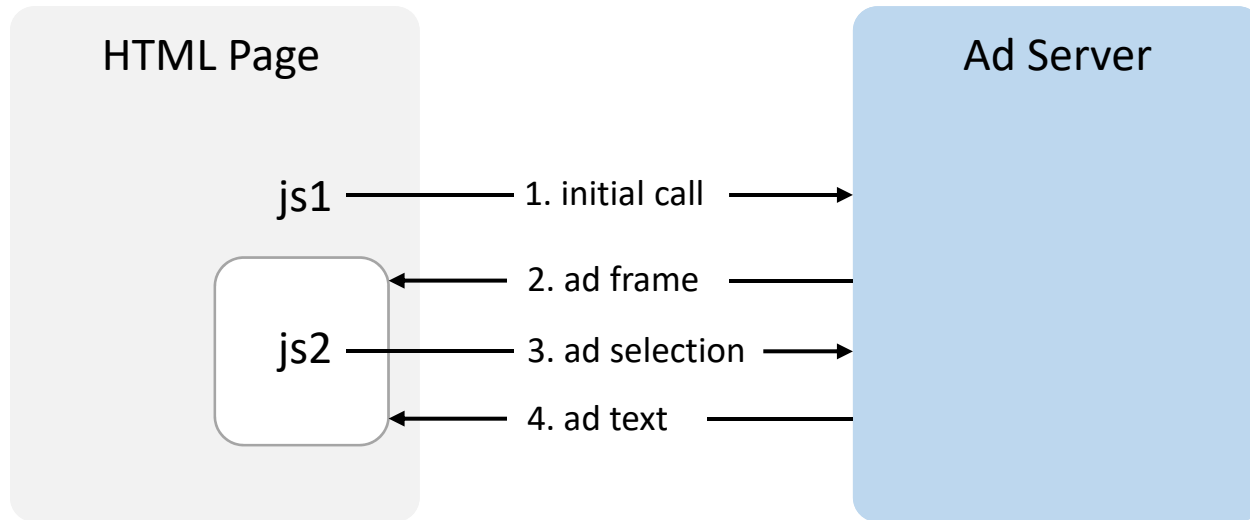
Bad Credit Guide
Looking to find bad credit? See our comprehensive guide.
Tradewestsales.Com

I Cured My Wrinkles
I Used (2) Secret Products To Get Rid Of My Wrinkles Forever.
www.GracesWrinkles.com

How Does it All Work: the Front End

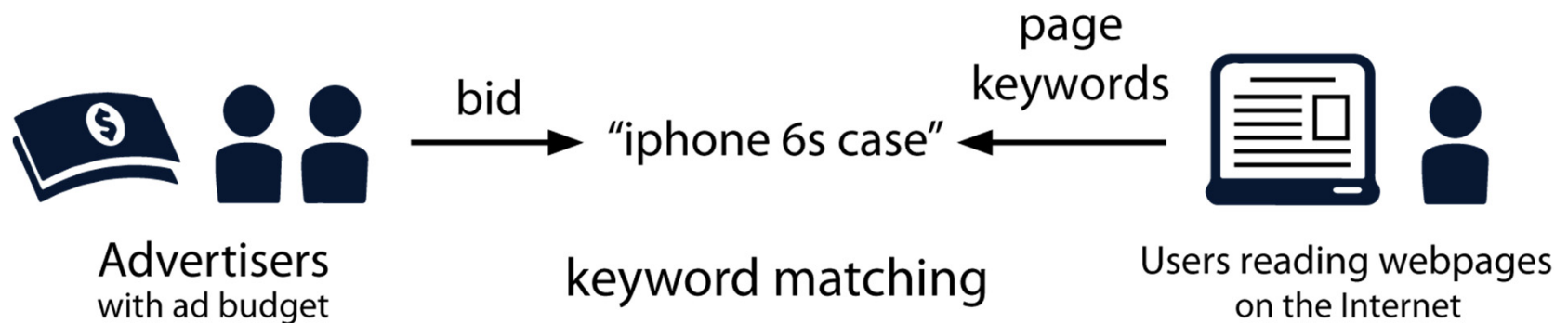
- Two main approaches

1. Page fully built by publisher using ads supplied by the ad network.
 - E.g.: XML feed (Usually done with large partners.)
2. Dynamic loading of ads



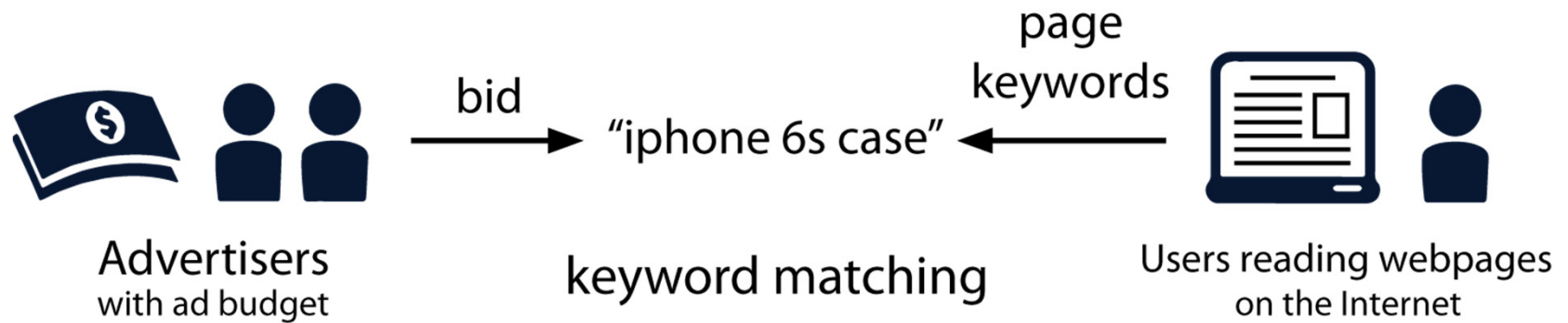
js: javascript s a high-level, dynamic, weakly typed, prototype-based, multi-paradigm, and interpreted programming language.

Contextual Ads and Sponsored Search



- Contextual ads is a natural way to expand the volume of textual ad campaigns in sponsored search
- No significant difference from advertisers' perspective
 - Same type of ads
 - Still bid keywords and participate keyword auctions
 - Can opt into contextual advertising for each ad group

Keywords Extraction for Contextual Ads



- Key technique in contextual ads is to extract good keywords from the page to raise the auction
 - The keywords still reveal the page topic (i.e., the user's intent)
 - The keywords might not necessarily occur in the page content

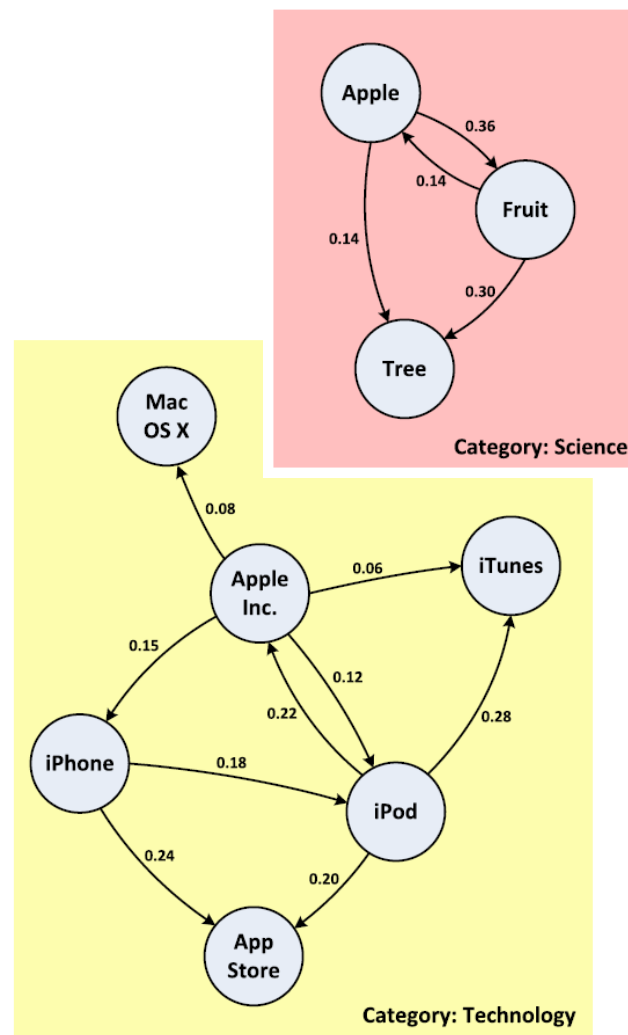
Techniques of Webpage Keyword Extraction (I)

- A machine learning approach
 - [Finding Advertising Keywords on Web Pages. Wen-tau Yih et al. WWW 2006.](#)
 - Human labels whether each candidate keyword in the page content is an appropriate keyword for this web page
 - Extract features for each candidate keyword in the page content
 - Train a machine learning model (e.g., logistic regression or SVM) with the data
 - Given a test web page, use the model to score the keywords and return the top-N results
 - Cons: (i) can only extract keywords in the web page;
(ii) cannot disambiguate the semantics of a keyword, e.g., apple

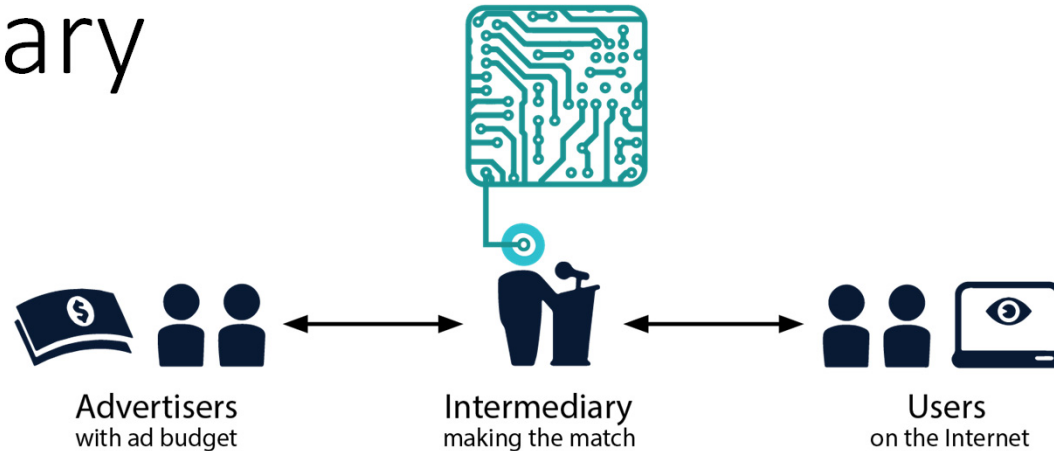
<i>features</i>	
A	all
-C	capitalization
-H	hypertext
-IR	IR
-Len	length
-Lin	linguistic
-Loc	location
-M	meta
-Ms	meta section
-Q	query log
-T	title
-U	URL

Techniques of Webpage Keyword Extraction (II)

- Two semantic approaches
 - A Semantic Approach to Contextual Advertising. Andrei Broder et al. SIGIR 2007.
 - Advertising Keywords Recommendation for Short-text Web Pages using Wikipedia. Weinan Zhang et al. TIST 2012.
- Leverage taxonomies or knowledge graphs to calculate semantic similarity between a keyword to a web page
- Node distance and PageRank algorithms to calculate the consensus of semantic topic of a keyword
 - E.g., apple for fruit or a company



Summary



- Computational advertising: design algorithms to make the best match between the advertisers and Internet users with economic constraints
- (Generalized) second price auction is the main type of trading ad display opportunities
- Search engine based sponsored search and contextual ads are the main types of CA before 2010
- Intelligent display advertising by real-time bidding and behavioral targeting emerges in 2011, which is focus of our next lecture