

WHENEVER I LEARN A NEW SKILL I CONCOCT ELABORATE FANTASY SCENARIOS WHERE IT LETS ME SAVE THE DAY.

OH NO! THE KILLER MUST HAVE FOLLOWED HER ON VACATION!



BUT TO FIND THEM WE'D HAVE TO SEARCH THROUGH 200 MB OF EMAILS LOOKING FOR SOMETHING FORMATTED LIKE AN ADDRESS!

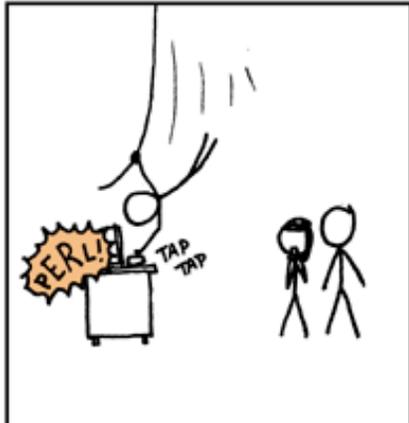


IT'S HOPELESS!

EVERYBODY STAND BACK.



I KNOW REGULAR EXPRESSIONS.





Online Advertising

David Kauchak

cs160

Fall 2009



Administrative

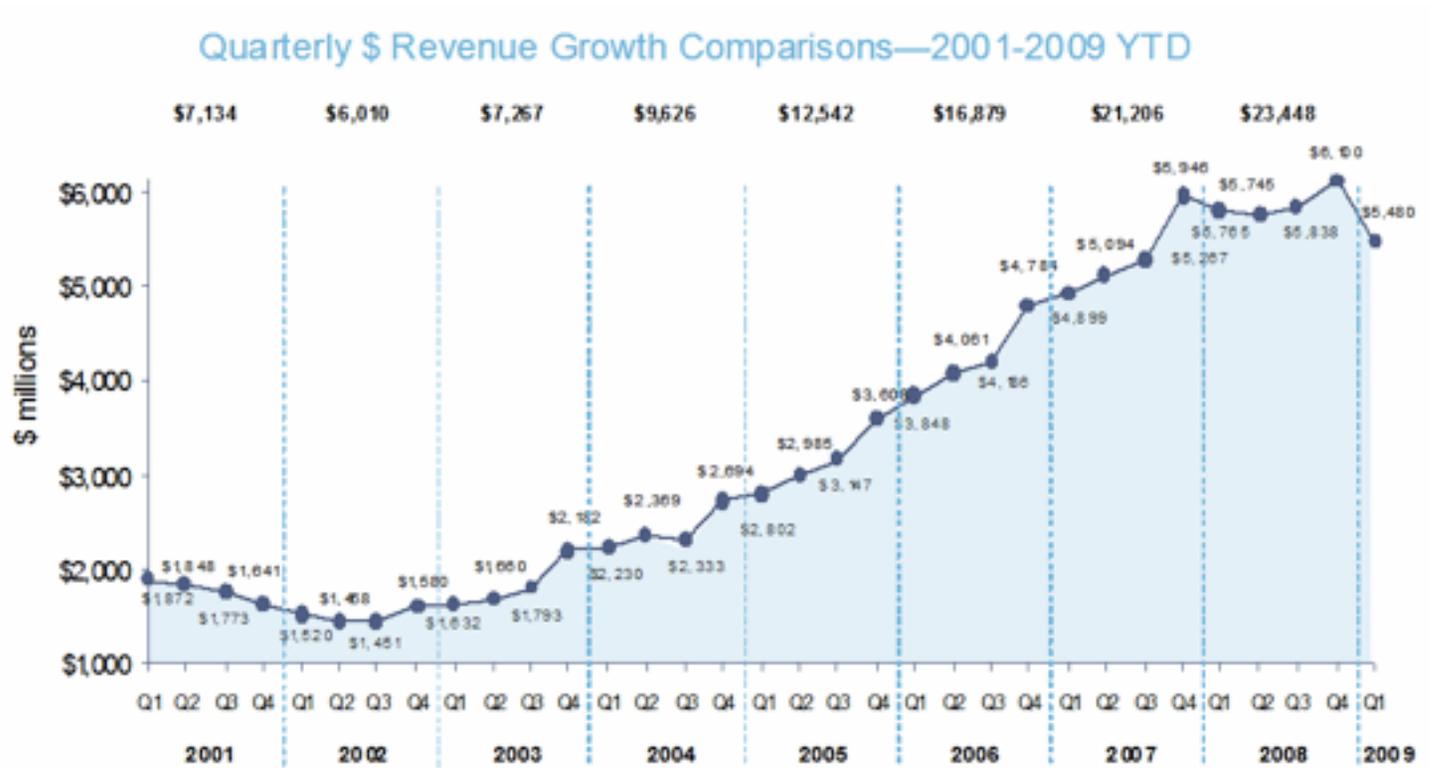
- CS Lunch Friday Frank West
 - Jeremy Frank (class of 1990)
- Project reports should be ~3 pages



SVN

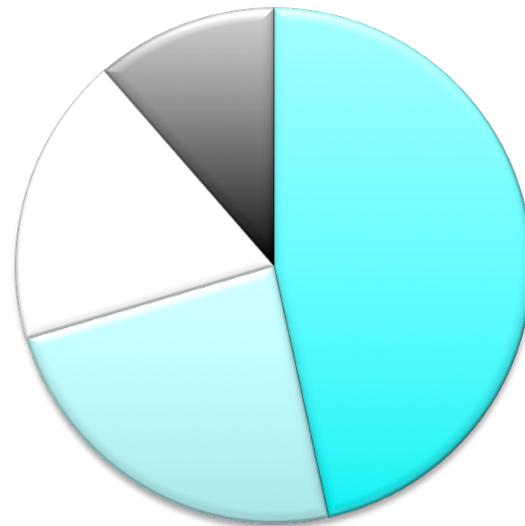
- Checkout the project
 - Run the following in your "src" directory of eclipse (though be careful not to overwrite your existing code!!!)
 - `svn checkout https://svn.cs.pomona.edu/cs160-f09/search`
 - refresh your eclipse project
- "svn add" to add new files (added files won't be added to the repository until you commit changes)
- "svn update" gets the latest version
- "svn diff" will get you the difference between your local file and the file(s) in the repository
- No changes are made to the repository until you commit your changes using "svn commit"
 - Only after you're sure that what you're going to commit compiles and works, should you commit your changes
- To get any of the data reading files with "DB" in the name, you'll need to link in the mysql...jar file into your build path

Online advertising \$



http://www.iab.net/about_the_iab/recent_press_releases/press_release_archive/press_release/pr-060509

Where the \$ comes from



- keyword search
- display
- classifieds
- other

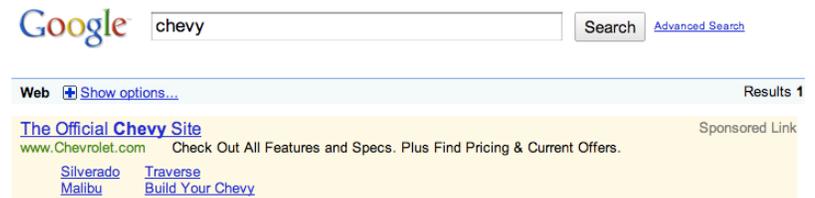
<http://www.informationweek.com/news/internet/reporting/showArticle.jhtml?articleID=207800456>

3 major types of online ads

- Banner ads



- Keyword linked ads



- Context linked ads



Banner ads

The screenshot shows the Yahoo! Autos website interface. At the top, there's a navigation bar with links like Home, New Cars, Used Cars, Research, Finance, Insurance, Maintain, and Green Center. Below this is a search bar and a 'Web Search' button. The main content area is divided into several sections: 'New Cars' with filters for Make and Model; 'Used Cars' with filters for Make and Model; 'Certified Pre-Owned Cars'; 'Car Finder' with filters for Price, Body Style, Driving and Performance, Fuel, and Seating and Capacity; 'Vehicle Spotlight' featuring 'Happy Honda Days!'; and 'Tips and Advice' featuring 'Secrets of a Professional Negotiator'. A large banner advertisement for GMC is positioned at the top of the main content area, featuring the text 'ALL NEW EQUINOX', '32 MPG HWY AND UP TO 600 MILES BETWEEN FILL-UPS', and '60-DAY SATISFACTION GUARANTEE'. Two red arrows point from the text 'standardized set of sizes' at the bottom to the banner and the '60-DAY SATISFACTION GUARANTEE' text.

standardized set of sizes



Ad formats

- Floating ad: An ad which moves across the user's screen or floats above the content.
- Expanding ad: An ad which changes size and which may alter the contents of the webpage.
- Polite ad: A method by which a large ad will be downloaded in smaller pieces to minimize the disruption of the content being viewed
- Wallpaper ad: An ad which changes the background of the page being viewed.
- Trick banner: A banner ad that looks like a dialog box with buttons. It simulates an error message or an alert.
- Pop-up: A new window which opens in front of the current one, displaying an advertisement, or entire webpage.
- Pop-under: Similar to a Pop-Up except that the window is loaded or sent behind the current window so that the user does not see it until they close one or more active windows.
- Video ad: similar to a banner ad, except that instead of a static or animated image, actual moving video clips are displayed.
- Map ad: text or graphics linked from, and appearing in or over, a location on an electronic map such as on Google Maps.
- Mobile ad: an SMS text or multi-media message sent to a cell phone.

Components for display advertising

User



Ad server



Advertiser



Publisher

The New York Times

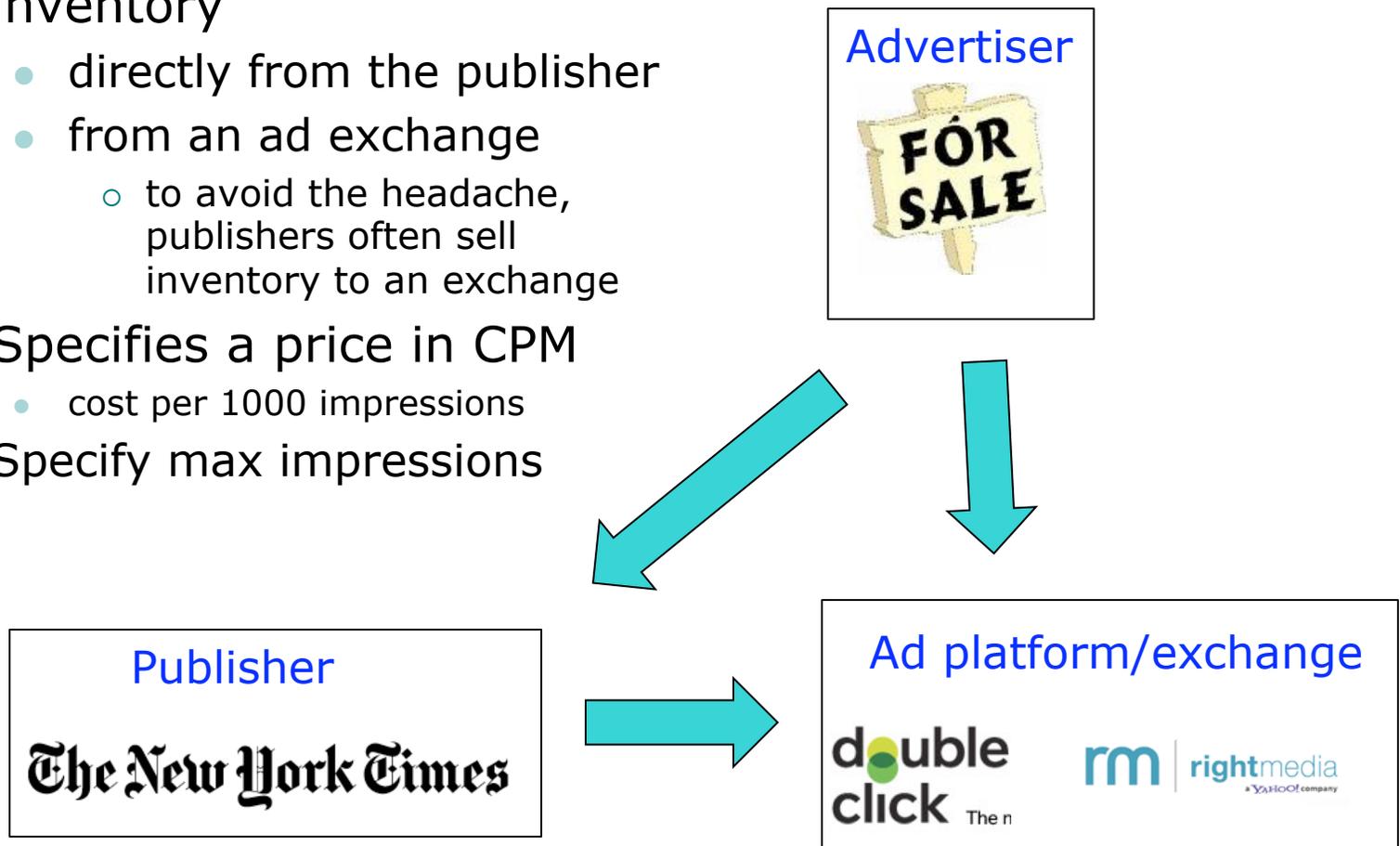
Ad platform/exchange

double
click The n

rm | **rightmedia**
a Yahoo! company

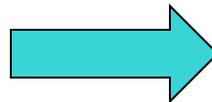
Banner ad process

- Advertiser “purchases inventory”
 - directly from the publisher
 - from an ad exchange
 - to avoid the headache, publishers often sell inventory to an exchange
- Specifies a price in CPM
 - cost per 1000 impressions
- Specify max impressions

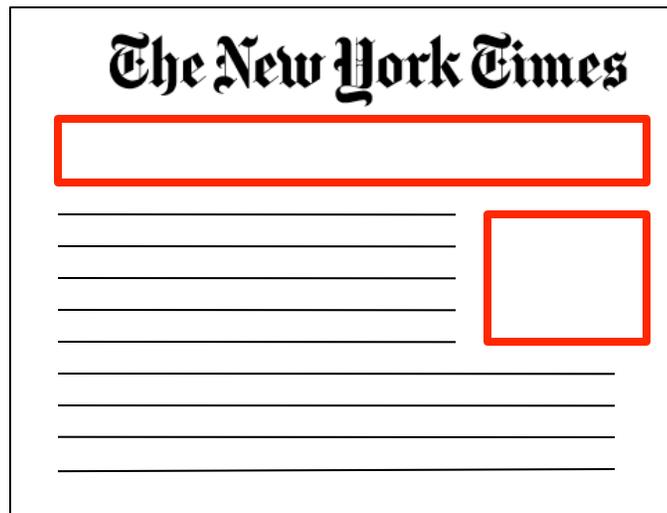
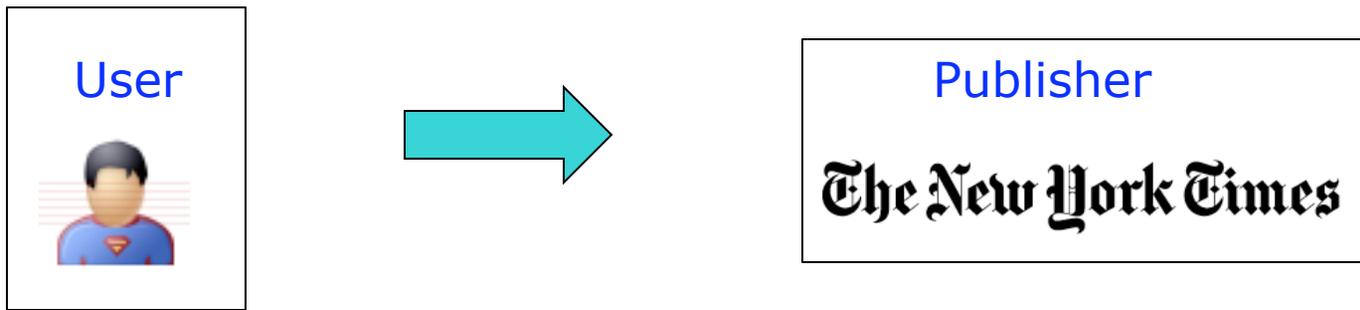


Banner ad process

- Advertiser uploads banners to banner server



Banner ad process



- User visits a page with places for ads
- Need to decide which ads to show

Banner ad process

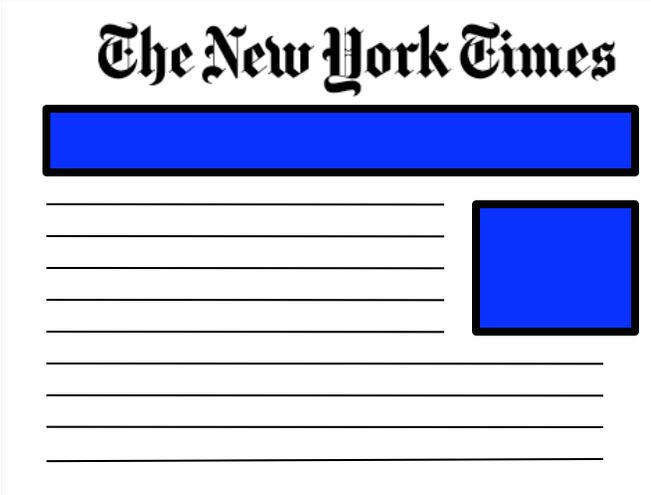
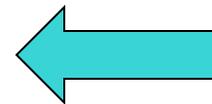
Publisher



Ad platform/exchange



Ad server





What are the problems/ inefficiencies with this process?

- Pricing
 - Fairly static: difficult to change price regularly
 - variable pricing based on user, etc
 - cpm pricing doesn't take into account clicks, revenue, etc.
- User targeting
 - We're only targeting users based on the site/page visited
 - What about a user that visits the same page everyday (e.g. nytimes)?
- Banner creation is fairly static
 - situation specific banners

User



Current trends: user targeting

- What information might we know about a user?
 - many of the sites a user has visited
 - cookies
 - everytime an ad is shown to a user, the ad is requested and we know which site the user is at
 - e.g. doubleclick cookie
 - Which ads the user has seen
 - Which ads the user has clicked on
 - Geographic information (via IP)
 - Demographic information (age, gender, profession, ...)
 - Signed in to Yahoo, Hotmail, etc.
 - Day of week, time of day, part of the month
 - Lots of other information
 - How much money they make
 - Whether they've bought anything recently
 - Mortgage payment
 - Habits, etc.



User targeting: RealAge

- Calculate your “biological age” based on a questionnaire
- 150 questions
- 27 million people have taken the test
- Information is used for marketing purposes

Health Interests

Choose the topics below that interest you and your family.

Medical Conditions

- ADD/ADHD, adult
- ADD/ADHD, child
- Allergies
- Arthritis, joint pain
- Asthma, adult
- Asthma, child
- Back pain
- Cancer
- Chronic pain
- Diabetes, type 2
- Diabetes, type 1
- Epilepsy
- Gastrointestinal conditions
- Headaches, migraines
- Heartburn
- Insomnia, sleep problems
- Osteoporosis, bone health

Women's Health

- Breast cancer
- Incontinence and overactive bladder
- Menopause
- Period problems
- Sexual health

Men's Health

- Prostate health
- Sexual health
- Hair loss

Cardiovascular Conditions

- Heart disease
- High blood pressure
- High cholesterol

Lifestyle

- Fitness and exercise
- Food and nutrition
- Healthy home
- Herbs and alternative medicine
- Living green
- Longevity and retirement
- Looking younger
- Parenting
- Reducing financial stress
- Reducing personal stress
- Weight management

Memory and Emotional Health

- Anxiety
- Bipolar disorder
- Depression
- Memory and mind

General Body Maintenance

- Dental care
- Foot care
- Healthy hair
- Hearing, ear care
- Skin care
- Vision, eye health

User targeting: data aggregation

- Companies aggregate this data
 - Bluekai
 - Excelate

Consumer search behaviors are aggregated at scale

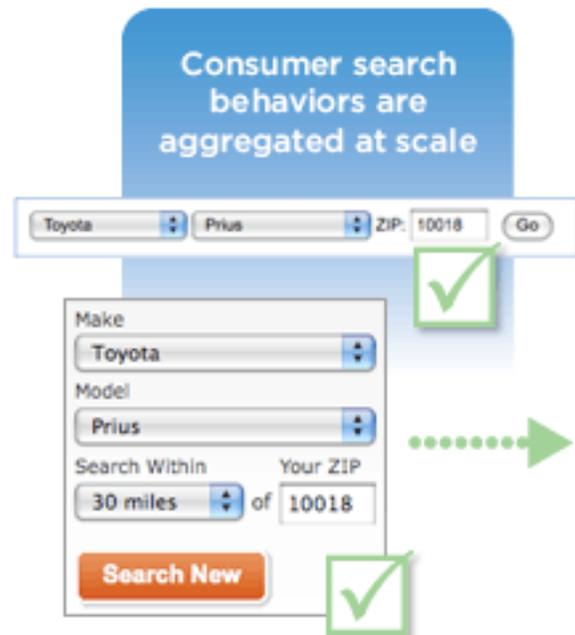
Toyota Prius ZIP: 10018 Go

Make
Toyota

Model
Prius

Search Within Your ZIP
30 miles of 10018

Search New



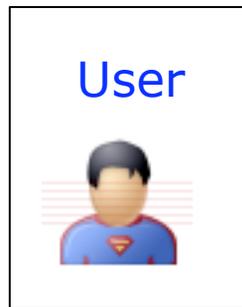
User targeting: Social networking sites



- Sites like myspace and facebook have lots of information about users, users' friends, etc
 - use content on a user's page
 - use information about a user's friends, e.g. purchases

User targeting: bottom line

- On a per impression basis, we have lots of information about the user the ad will be shown to



age
gender
location
income
search history
number of ad views
...



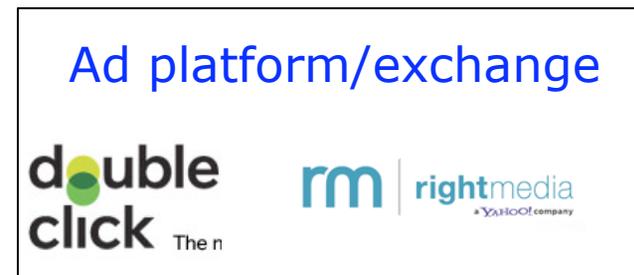
Banner ad pricing

- Advertising exchange
 - Auction-based system for purchasing ads
 - Auction happens roughly per impression
 - Auction targeting based on user characteristics
 - recent trend (last year or two)

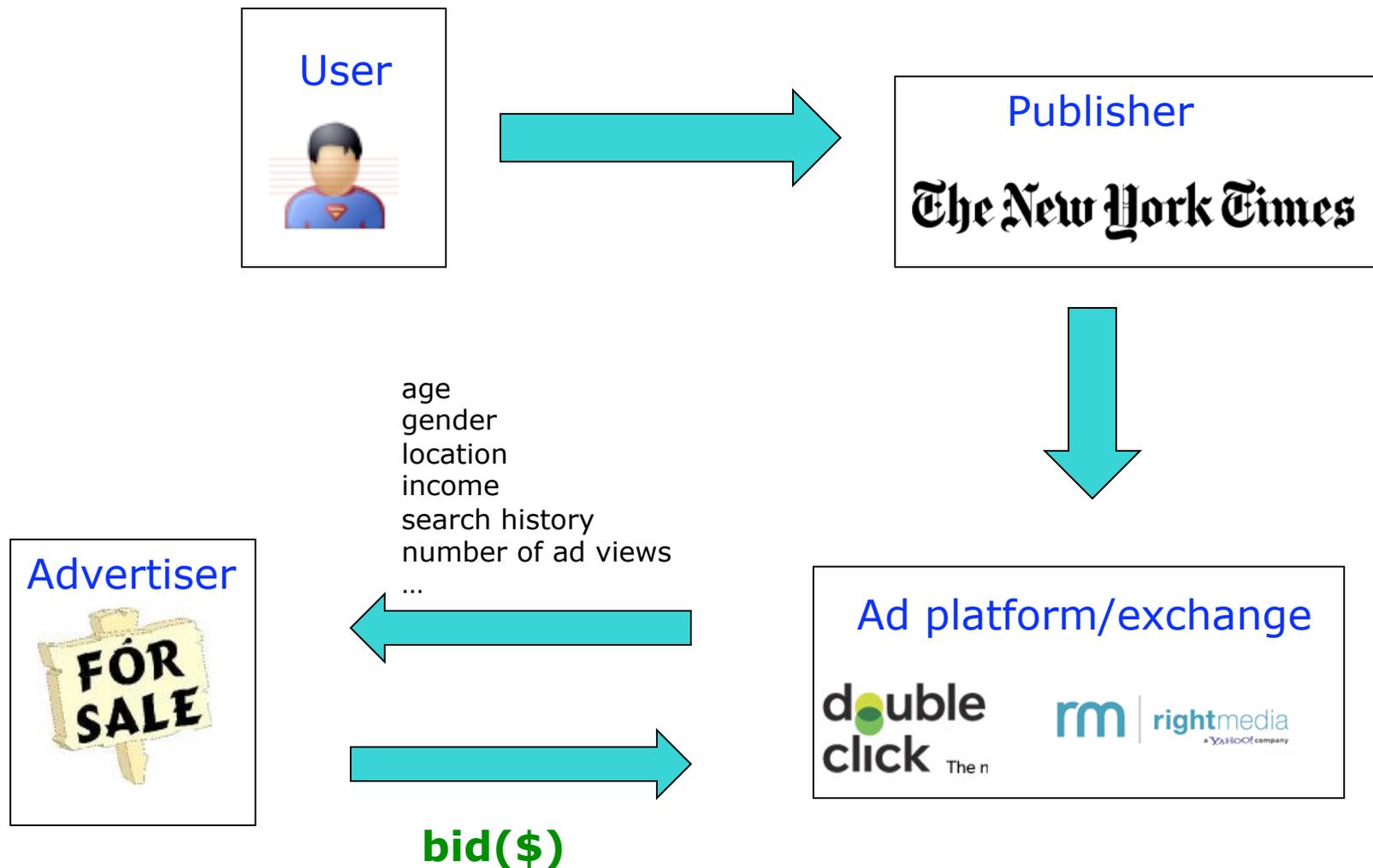
\$3 CPM for men, ages 20-25, CA NY FL from 12-5pm

Banner ad exchanges

- Advertiser “uploads” bids to exchange
 - via spreadsheet
 - or programmatically
 - Specify targeting
 - Can also set thresholds on user views
- Auction is performed by exchange
- **Downsides?**



Ideal ad exchange: true auction



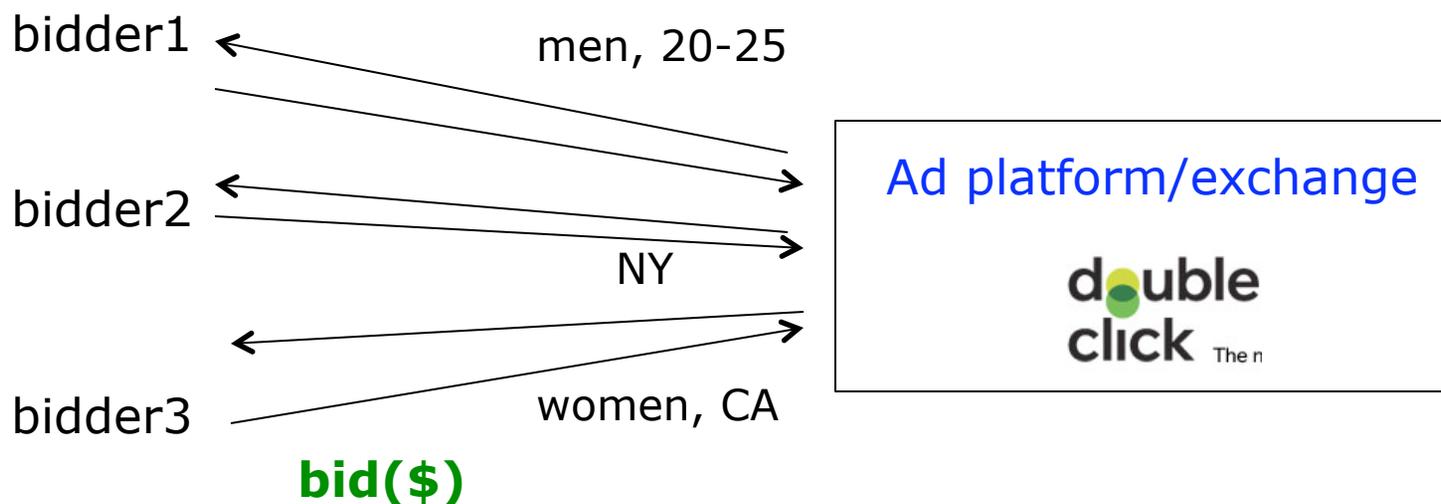


True auction: technical challenges

- We need to make a decision quickly (on the order of a few hundred ms)
 - multiple advertisers
 - advertiser must make decision
 - network latency
 - perform auction
 - this happens millions of times a day
 - ...

True auction: some first attempts

- Doubleclick "callback"
 - specify a "bidder" based on some targeting specifications
 - bidder only bids on impressions that match criterion





True auction: AppNexus

- Ex-RightMedia folks
- Initially, cloud computing
- Advertiser runs a bidder server side
 - avoid network latency
 - auction is self-contained at the exchange
 - Requires framework on exchange side for security, speed, etc.



Pricing

- Advertisers don't care about CPM
 - CPC (cost per click)
 - CPA (cost per action)
 - RPM (revenue per impression)
- Some work to move exchanges towards this
- **Challenge?**
 - Need to estimate these from data
 - Data is very sparse $\sim 1/1000$ people click
 - Similar order of magnitude purchase (though depends on the space)

Paid search components

User



Advertiser



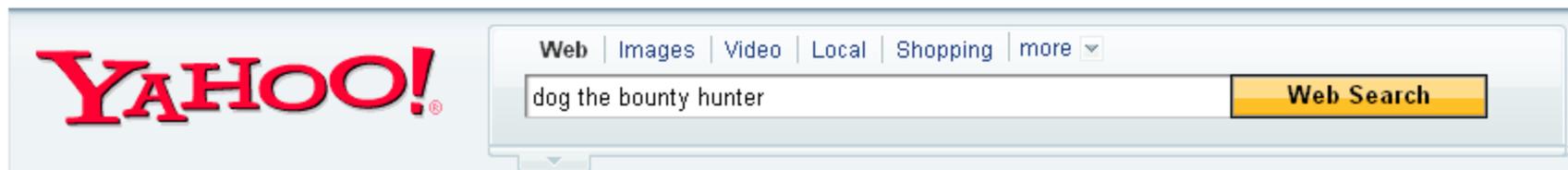
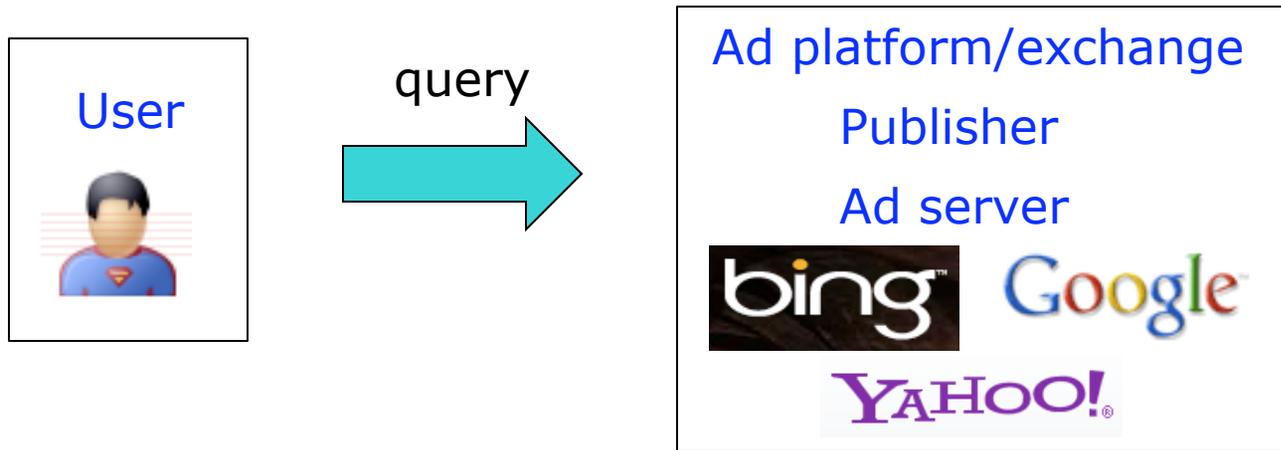
Ad platform/exchange

Publisher

Ad server



Paid search query



Web Images Video Local Shopping more

dog the bounty hunter

Search

Options

YAHOO!

1-10 of 3,190,000 for dog the bounty hunter (About) - 0.12 sec

Also try: [dog the bounty hunter costume](#), [dog the bounty hunter daughter](#), [More...](#)

SPONSOR RESULTS

- [Barkley Square](#)
[www.barkleysquare.net](#) - Spoil your pet with our Pooch Paradise.
- [How to Be a Dog Bounty Hunter](#)
[www.search4degrees.com](#) - Law Enforcement & Criminal Justice Degrees Online - Request Info Now.

1. [DOG THE BOUNTY HUNTER](#)

Official site of the famous **bounty hunter** Duane "Dog" Chapman, known for numerous captures including that of convicted rapist and Max Factor heir Andrew Luster.
[www.dogthebountyhunter.com](#) - 7k - [Cached](#)

2. [Dog the Bounty Hunter on A&E TV](#)

Reality series that follows the adventures of **bounty hunter** Duane **Dog** Chapman.
[www.aetv.com/dog_the_bounty_hunter/index.jsp](#) - 43k - [Cached](#)

3. [Dog the Bounty Hunter - Wikipedia, the free encyclopedia](#)

You can't run, but you can hide (up high) from **Dog the Bounty Hunter** " ... **Bounty Hunters** Assail Duane 'Dog' Chapman - Associated Press story ...
Quick Links: [2006 arrest and extradition proceedings](#) - [In popular culture](#) - [Trivia](#)
[en.wikipedia.org/wiki/Dog_the_Bounty_Hunter](#) - 39k - [Cached](#)

4. [Dog the Bounty Hunter - IMDb](#)

SPONSOR RESULTS

[Become a Bounty Hunter](#)

Find out how. Request fast, free Criminal Justice Degree information.
[www.criminal-justice-online.com](#)

[Take a Free Career Test](#)

Match Jobs To Your Personality. A Serious Tool To Help Guide You.
[Free-Career-Test.com](#)

[Dog the Bounty Hunter Costume](#)

Find **dog the bounty hunter** costume at Great Prices.
[www.Pronto.com](#)

[How to Become a Bounty Hunter](#)

Get educated to fight crime with a degree in Criminal Justice online.
[www.AIU-DegreesOnline.com](#)

[Monster Jam Diecast Trucks Hot Wheels](#)

Grave Digger, Superman, The Destroyer, Annihilator and More

What is required of the advertiser?



A bit more structure than this...



millions of keywords

<100K keywords

campaign1

...

<100 keywords

adgroup1

adgroup2

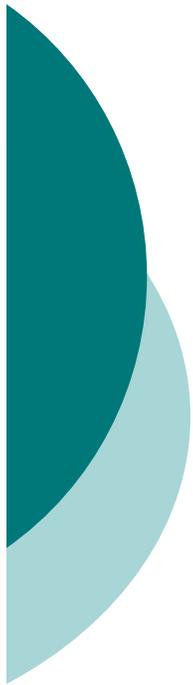
adgroup3

...

keyword1

keyword2

...





Adgroups

- Adgroups are the key structure
- Adcopy and landing pages are associated at the adcopy level
- Keywords should be tightly themed
 - promotes targeting
 - makes google, yahoo, etc. happy

Creating an AdWords Ad



[Help](#) | [Sign out](#)

Jump to previous customer...

Open tool...

[Advanced Search](#)

Ops Control Center

Campaign Management

Reports

Analytics

My Account

[Campaign Summary](#) | [Tools](#) | [Conversion Tracking](#)

Search my campaigns:

[Campaign Summary](#) > [Seattle Condo](#) > **Ad Group #1**

1 of 2 Ad Group(s) [Next >](#)

Ad Group: Ad Group #1 5204121

[Seattle View Penthouse](#)
 2 Bed, 2 Bath in Fantastic Location
 Mountain Views, Huge private deck
www.badros.com/view-condo-fsbo.html

Paused

[View History: this ad group](#)

Ad Group Approval Bin : [Primary](#) | [Secondary](#) | [All](#)

1 of 1 - [view all](#)

FamilySafe

Summary

Keywords

Ad Variations

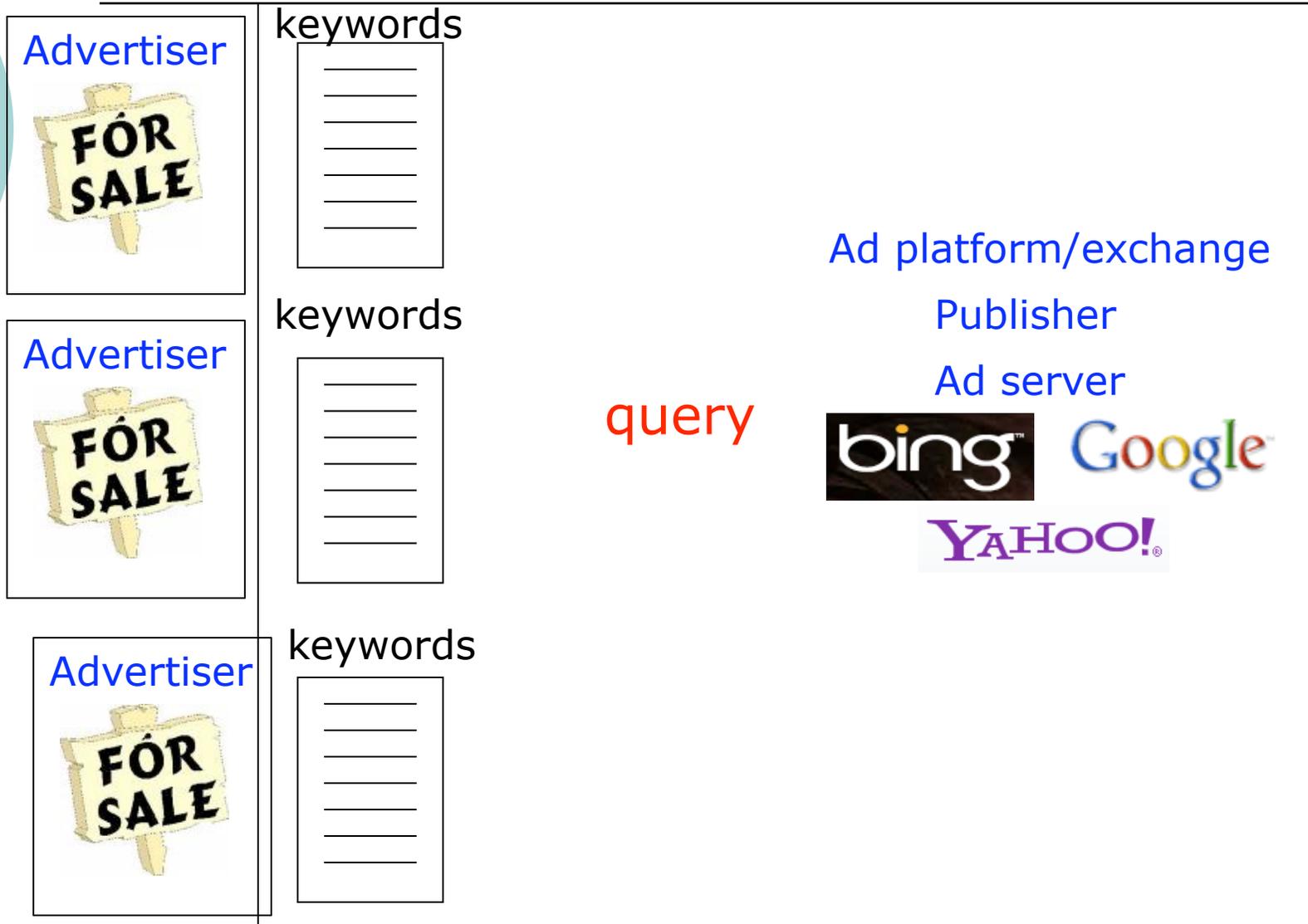
Feb 14, 2003 to Jan 20, 2006 [Change range](#)

+ Add keywords: [Quick add](#) | [Keyword tool](#) [Edit Keywords](#) | [Search this list](#)

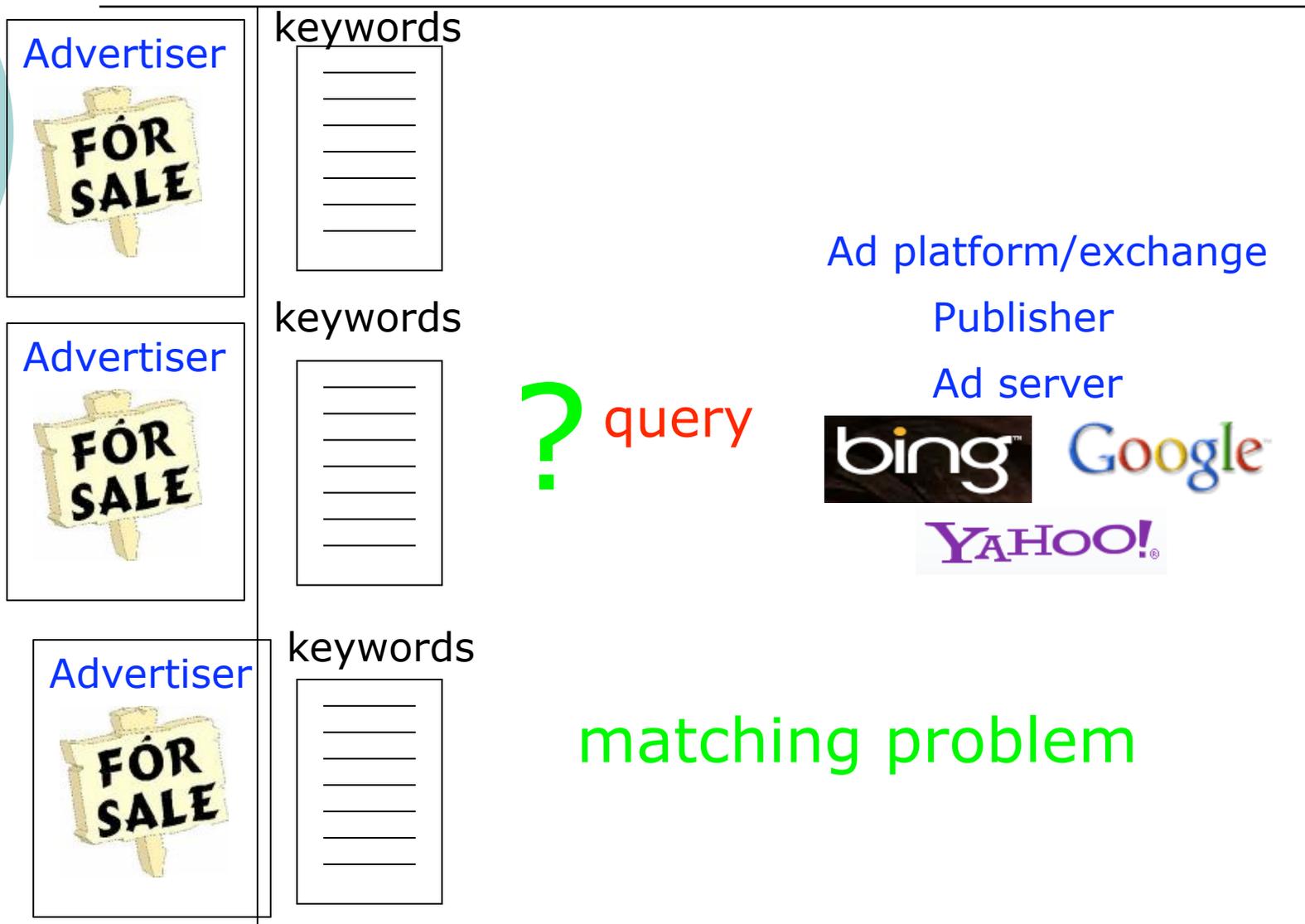
1 - 11 of 11 keywords.

<input type="checkbox"/> Keyword	Status [?]	Current Bid <input type="text" value="Max CPC"/>	Clicks	Impr.	CTR	Avg. CPC	Cost	Avg. Pos
Total	Enabled	Default \$1.00 [edit]	456	22,864	1.99%	\$0.46	\$209.47	3.0
<input type="checkbox"/> seattle apartment	Active	\$1.00	125	5,634	2.21%	\$0.30	\$37.26	1.1
<input type="checkbox"/> seattle condo	Active	\$1.00	143	2,906	4.92%	\$0.43	\$61.89	1.2
<input type="checkbox"/> seattle condominium	Active	\$1.00	50	1,296	3.85%	\$0.33	\$16.54	1.3
<input type="checkbox"/> seattle fremont apartment	Active	\$1.00	2	79	2.53%	\$0.06	\$0.12	1.1
<input type="checkbox"/> seattle fremont real estate	Active	\$1.00	1	53	1.88%	\$0.81	\$0.81	4.0
<input type="checkbox"/> seattle fremont condo	Active	\$1.00	2	47	4.25%	\$0.28	\$0.57	1.2
<input type="checkbox"/> seattle luxury apartment	Active	\$1.00	2	24	8.33%	\$0.09	\$0.17	1.1
<input type="checkbox"/> seattle fremont house	Active	\$1.00	1	22	4.54%	\$0.54	\$0.54	1.9
<input type="checkbox"/> seattle luxury condo	Active	\$1.00	4	19	21.05%	\$0.55	\$2.19	1.5
<input type="checkbox"/> seattle luxury condominium	Active	\$1.00	1	7	14.28%	\$0.18	\$0.18	2.0

Behind the scenes

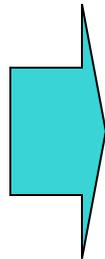
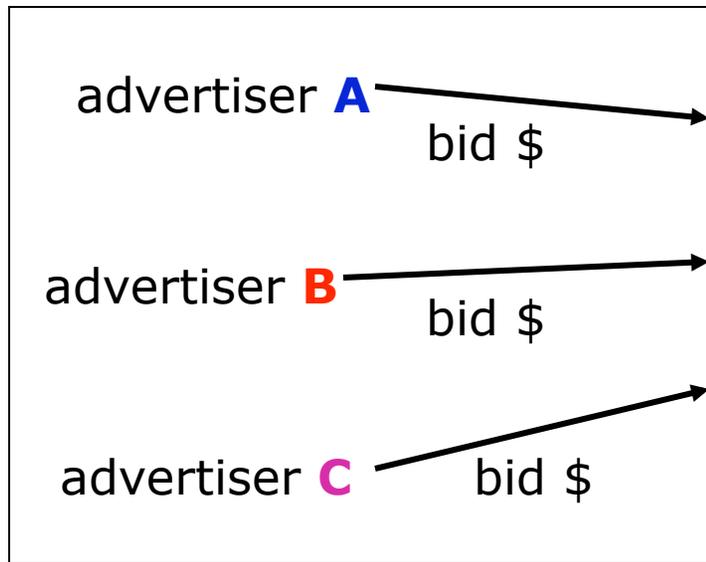


Behind the scenes

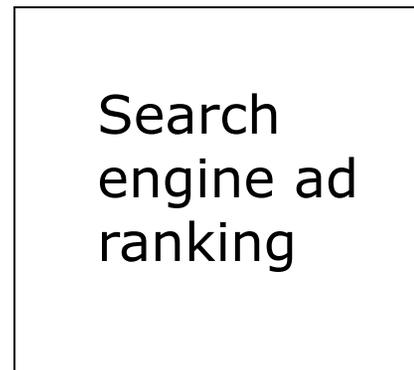
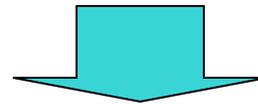


Behind the scenes

For all the matches...

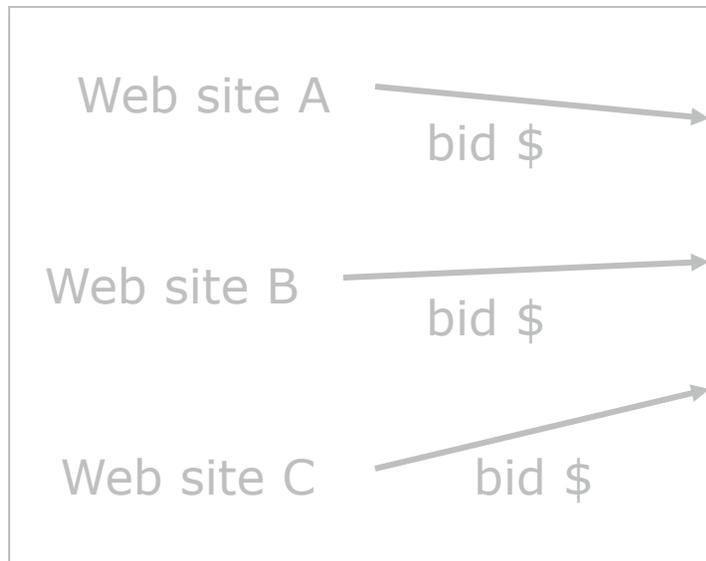


Other data (site content, ad content, account, ...)

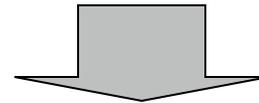


Behind the scenes: keyword auction

Site bids for keyword:
"dog the bounty hunter"



Other data (site
content, ad content,
account, ...)



Display
ranking

Web site **B**

Web site **A**

Web site **C**



Search ad ranking

- Bids are CPC (cost per click)
- How do you think Google determines ad ranking?

$$\text{score} = \underbrace{\text{CPC} * \text{CTR}} * \underbrace{\text{“quality score”}} * \underbrace{\text{randomness}}$$

cost/clicks * clicks/impression =
cost/impression

Is it a good web pages?
Good adcopy?
Adcopy related to keyword?

Enhances user experience,
promoting return users

don't want people
reverse engineering
the system

data gathering



1st price auction

- Each bidder pays what they bid
- Not used by search engines. Why?
- Don't work well for repeat auctions!

Bidder	value	Bid1
A	9	10
B	7	5

How would the bids change next time (assuming a blind auction)?



1st price auction

- Each bidder pays what they bid
- Not used by search engines. Why?
- Don't work well for repeat auctions!

Bidder	value	Bid1	Bid2
A	9	10	7
B	7	5	6

A is going to want to decrease it's bid
B increase



1st price auction

- Each bidder pays what they bid
- Not used by search engines. Why?
- Don't work well for repeat auctions!

Bidder	value	Bid1	Bid2	Bid3
A	9	10	7	6
B	7	5	6	7

A decrease
B increase



1st price auction

- Each bidder pays what they bid
- Not used by search engines. Why?
- Don't work well for repeat auctions!

Bidder	value	Bid1	Bid2	Bid3	Bid4
A	9	10	7	6	8
B	7	5	6	7	7



1st price auction

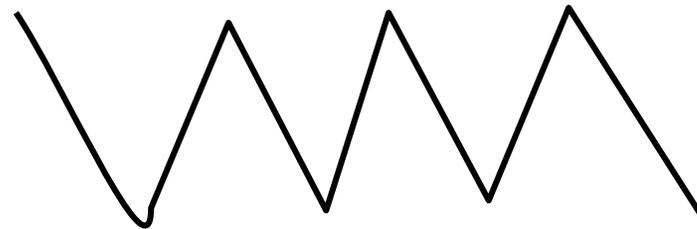
- Each bidder pays what they bid
- Not used by search engines. Why?
- Don't work well for repeat auctions!

Bidder	value	Bid1	Bid2	Bid3	Bid4	Bid5
A	9	10	7	6	8	8
B	7	5	6	7	7	5

1st price auction

- Each bidder pays what they bid
- Not used by search engines. Why?
- Don't work well for repeat auctions!

In general, tend to end up with unstable bids in a "sawtooth" pattern

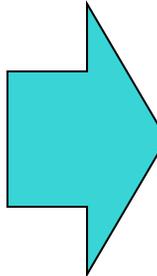


- bid down when you're winning
- bid up to get back in first
- bid back down

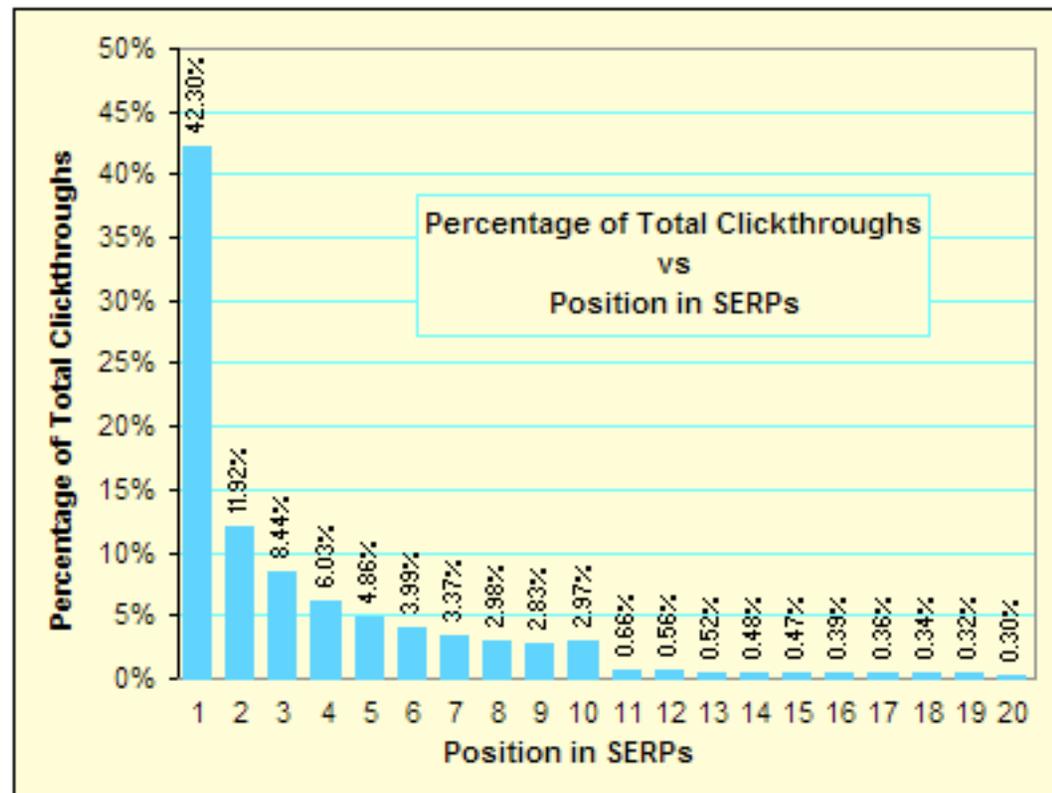


Auction system

- 2nd price auction
 - Winner pays one penny more than the 2nd place bid
 - Slightly complicated by modified scoring
 - Avoids sawtooth problem, but still not perfect

Bidder	Bid		Bidder	Price
A	10		A	5.01
B	5		B	1.01
C	1		C	1

CTR with respect to position



Note, these are not CTRs, but relative CTRs



Predicted CTR

score = CPC * CTR * "quality score" * randomness

- Any problem with using CTR of a keyword?
 - Zipf's law: most keywords get very little traffic
 - CTRs are generally $\sim 1-3\%$
 - Need a lot of impressions to accurately predict CTR
 - New advertisers, new adcopy, ...
- Major prediction task
 - machine learning
 - lots of features
 - share data within an advertiser and across advertisers

Factors affecting revenue for search engine

Monetization
(RPM)

$$\begin{aligned}
 &= \frac{\text{Revenue}}{\text{Queries}} \\
 &= \frac{\text{Revenue}}{\text{Clicks}} \times \frac{\text{Clicks}}{\text{Queries}} \\
 &= \frac{\text{Revenue}}{\text{Clicks}} \times \frac{\text{Queries w/ Ads}}{\text{Queries}} \times \frac{\text{Ads}}{\text{Queries w/ Ads}} \times \frac{\text{Clicks}}{\text{Ads}} \\
 &= \underbrace{\text{CPC}} \times \underbrace{\text{Coverage} \times \text{Depth}} \times \underbrace{\text{CTR per Ad}}
 \end{aligned}$$

Price
Quantity
Quality



Increasing search engine revenue

- Increase CPC (cost per click)
 - Increase conversion rate (i.e. post click performance)
 - Increase competition (higher bids)
- Increase coverage and depth
 - More keywords
 - more keywords per advertiser (i.e. keyword tools)
 - more advertisers
 - More broadly matching keywords to queries
- Increase CTR (click through rate)
 - Show more relevant ads in higher positions
 - Encourage high quality ads
 - Precise keyword/query matching



Advertiser margin

margin = revenue - cost

$$= \frac{\text{Revenue}}{\text{Action}} \times \frac{\text{Actions}}{\text{Impression}} \times \text{Impressions} - \text{cost}$$

$$= \frac{\text{Revenue}}{\text{Action}} \times \frac{\text{Actions}}{\text{Click}} \times \frac{\text{Clicks}}{\text{Impression}} \times \text{Impressions} - \text{cost}$$

$$= \frac{\text{Revenue}}{\text{Action}} \times \frac{\text{Actions}}{\text{Click}} \times \frac{\text{Clicks}}{\text{Impression}} \times \text{Impressions} - \frac{\text{Cost}}{\text{Click}}$$

revenue per
transaction

conversion
rate

CTR

CPC



Increasing advertiser margin

- Increase revenue per transaction
 - sales, marketing
 - increase price
- Increase conversion rate (actions per click)
 - better landing page
 - better offers
 - cheaper price
 - more offers/options
- Increase click through rate
 - better adcopy
- Increase impressions
 - more keywords
- Decrease cost per click
 - decrease bid
 - increase "quality score"
 - bid on less competitive keywords

Contextual advertising



c h e e s e
www.cheese.com

Home

Get Cheese News By Email

Subscribe

- Cheese by names
- Cheese by countries
- Cheese by textures
- Cheese by milk
- Vegetarian cheeses
- Alphabetical list of cheeses
- Cheese facts
- Suppliers
- Cheese recipes
- Bookstore
- Links
- Headlines

Welcome to the #1 resource for cheese!
It's all about cheese!

CHEESE SEARCH:
Enter cheese name in the box below. Don't worry about spelling.
 Search!
There are currently 654 cheeses in our database.

Ads by Google [English Cheese](#) [Cheese Stilton](#) [Buffalo Cheese](#) [Goats Cheese](#)

Cheese is nutritious food made mostly from the milk of cows but also other mammals, including sheep, goats, buffalo, reindeer, camels and yaks. Around 4000 years ago people have started to breed animals and process their milk. That's when the cheese was born.

Explore this site to find out about different kinds of cheeses from around the world.

You can search the database of 654 cheeses by names, by country of origin, by kind of milk that is used to produce it, or by texture.

The database includes information on most famous cheeses such as Cheddar, Camembert, Stilton or Parmesan, as well as rarities like Crotin du Chavignol.

We are planning to add a new section on recipes. Do you know a good one? Please, send it to us and we will publish it with a credit to you!

Do not hesitate to contact us if you have any comments on this site! We will do our best to make this site fit better your needs and expectations.

Recipes
Get free recipes, plan meals, print coupons & more at Albertsons online
[Ads by Google](#)

Lose 20 lbs in 3 Weeks
Amazing Chinese fat-loss secret. As seen on Oprah & 60 Minutes.
[Advertise on this site](#)

Gourmet Cheese Recipes
Add satiny ripe brie or tangy Goat cheese to your next dish
[www.iledefrancecheese.](#)

Wisconsin Cheese Mart
Since 1938. The World's Largest Selection of Wisconsin Cheese.
[www.WisconsinCheeseM](#)

Free English Cheddar
Receive Free Farmhouse Cheddar England's Gourmet Gem Survey Now
[EnglishCheddar.food-off](#)

Goat Milk Cheese
Indulging Goat Milk Cheese Free S&H, Great Tasting
[www.cheesesupply.com/](#)

[Advertise on this site](#)

Cheese delivered to your door





Contextual Advertising

- Text ads on web pages
- Uses similar technology and framework to search advertising
 - Advertiser supplies keywords, adgroups, adcopy, bids
 - Rather than match queries, match text on page
- Some differences
 - A lot more text, so many more matches and multiple matches
 - Generally lower CTRs, lower conversion performance, adjustments made in payment
- Easy way for search engines to expand revenue
- Challenges
 - extracting “keywords” from a web page
 - be careful about matching. e.g. wouldn’t want to show a competitors ad

How the ads are served

```
function google_show_ad() {  
  var w = window;  
  w.google_ad_url = 'http://pagead2.googlesyndication.com/pagead/ads?' +  
    '&url=' + escape(w.google_page_url) +  
    '&hl=' + w.google_language;  
  document.write('<ifr' + 'ame' +  
    ' width=' + w.google_ad_width +  
    ' height=' + w.google_ad_height +  
    ' scrolling=no></ifr' + 'ame>');  
}  
google_show_ad();
```



Lots of problems in online advertising

- Display (banner ads)
 - Banners on the fly
 - User targeting
 - Predict performance based on user data
 - Tracking users
 - auctions
 - buyer strategy
 - auction holder policies
 - Banner/ad selection



Lots of problems in online advertising

- Paid search
 - keyword generation
 - adgroup generation
 - keyword performance estimation
 - impressions/volume, CTR, conversion rate, rev.
 - adcopy generation
 - bid management
 - auction mechanisms
 - keyword/query matching



Lots of problems

- Misc

- Data analysis
 - What works well
 - Trends in the data
 - Anomalies
- click fraud
- scale (many of these things must happen fast!)
- Landing page optimization



Typical CPMs in advertising

- Outdoor: \$1-5 CPM
- Cable TV: \$5-8 CPM
- Radio: \$8 CPM
- Online
 - Display \$5-30 CPM
 - Contextual: \$1-\$5 CPM
 - **Search: \$1 to \$200 CPM**
- Network/Local TV: \$20 CPM
- Magazine: \$10-30 CPM
- Newspaper: \$30-35 CPM
- Direct Mail: \$250 CPM