
Internet Advertising
Online Advertising
A cross-discipline of
information retrieval, machine learning, and economics
Microsoft Research

The Speaker

- Tie-Yan Liu
  - Lead Researcher, Microsoft Research Asia
  - Co-author of 70+ papers in SIGIR, WWW, NIPS, ICML, KDD, etc; co-author of SIGIR Best Student Paper (2008) and JVCIR Most Cited Paper Award (2004~2006); author of two books (Now publisher and Springer).
  - Associate editor, ACM Transactions on Information System; Editorial board member, IR Journal and ISRN Artificial Intelligence.
  - PC Chair of RIAO 2010, Area Chair of SIGIR (2008~2011), Track Chair of WWW (2011).
  - Senior member of the IEEE and the CCF.

Advertising

- Advertising is a form of communication intended to persuade an audience (viewers, readers, or listeners) to purchase or take action upon products, ideals, or services.
- Venues
  - Magazines,
  - Billboards,
  - Newspapers,
  - Handbills,
  - TV, etc.

Advertising

- Advertisers are buying attention!
  - Advertisers want the attention of certain people
  - People are only open to certain ads

    (Herbert Simon, 1971)
- "Half the money I spend on advertising is wasted; the trouble is I don't know which half."

  (John Wanamaker)

Advertising is Not Easy!

Online Advertising

- Advertising in the context of online services
  - Users: online users with search and browsing behaviors
  - Publishers: search engines and websites
  - Advertisements: ads copies and landing pages

Online Advertising

- First online advertisement
  - In 1994. 10. 24.
  - Place on the website of HotWired.
  - Created for Modem Media/AT&T by TANGENT Design/Communications of Westport, CT.
  - A staggering CTR of 42%!

Why Online

- Classical:
  - High cost per venue ($3Mil for a Super Bowl TV ad)
  - No personalization
  - Targeting by the wisdom of people
  - Hard to measure ROI
- Online (almost the exact opposite):
  - Billions of opportunities and tiny cost per opportunity
  - Personalizable
  - Targeting by algorithms
  - Much more quantifiable (Impression, CTR, CPC, …)

Online Advertising vs. Offline Advertising

Advertising Spending Worldwide, by Media, 2007-2011 (millions)

	2007	2008	2009	2010	2011
TV	$180,460	$185,788	$172,320	$174,836	$183,177
Newspapers	$130,178	$123,109	$102,136	$97,703	$97,228
Internet	$40,242	$49,544	$54,087	$60,253	$68,557
Magazines	$59,196	$56,588	$45,415	$42,762	$42,573
Radio	$38,583	$37,630	$33,647	$33,280	$34,216
Outdoor	$31,752	$31,888	$29,112	$29,828	$31,430
Cinema	$2,268	$2,377	$2,180	$2,274	$2,422
Total	**$482,680**	**$486,924**	**$438,896**	**$440,936**	**$459,603**

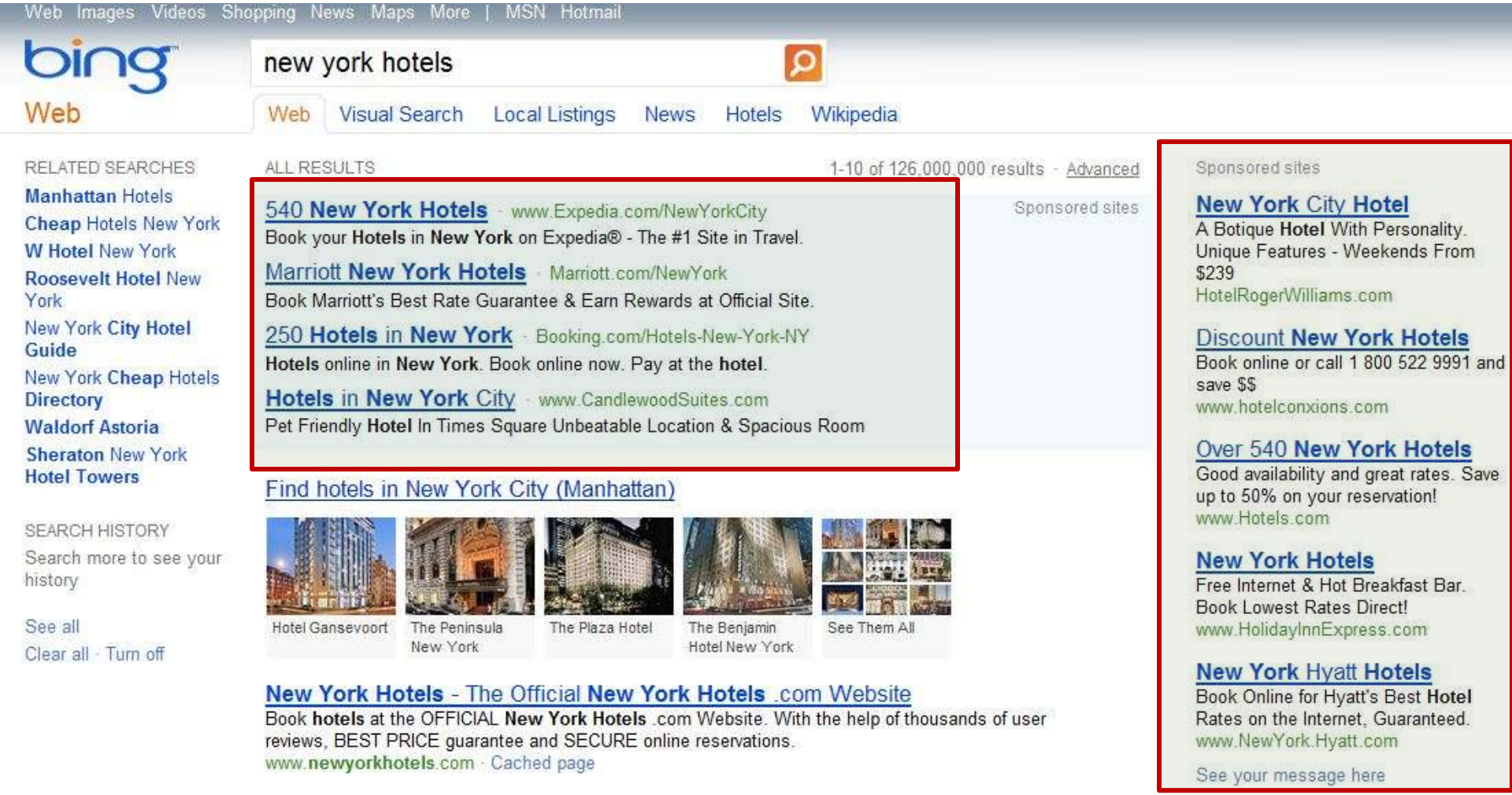
Note: currency conversion at 2008 average rates
Source: ZenithOptimedia as cited in press release, October 19, 2009

107790 www.eMarketer.com

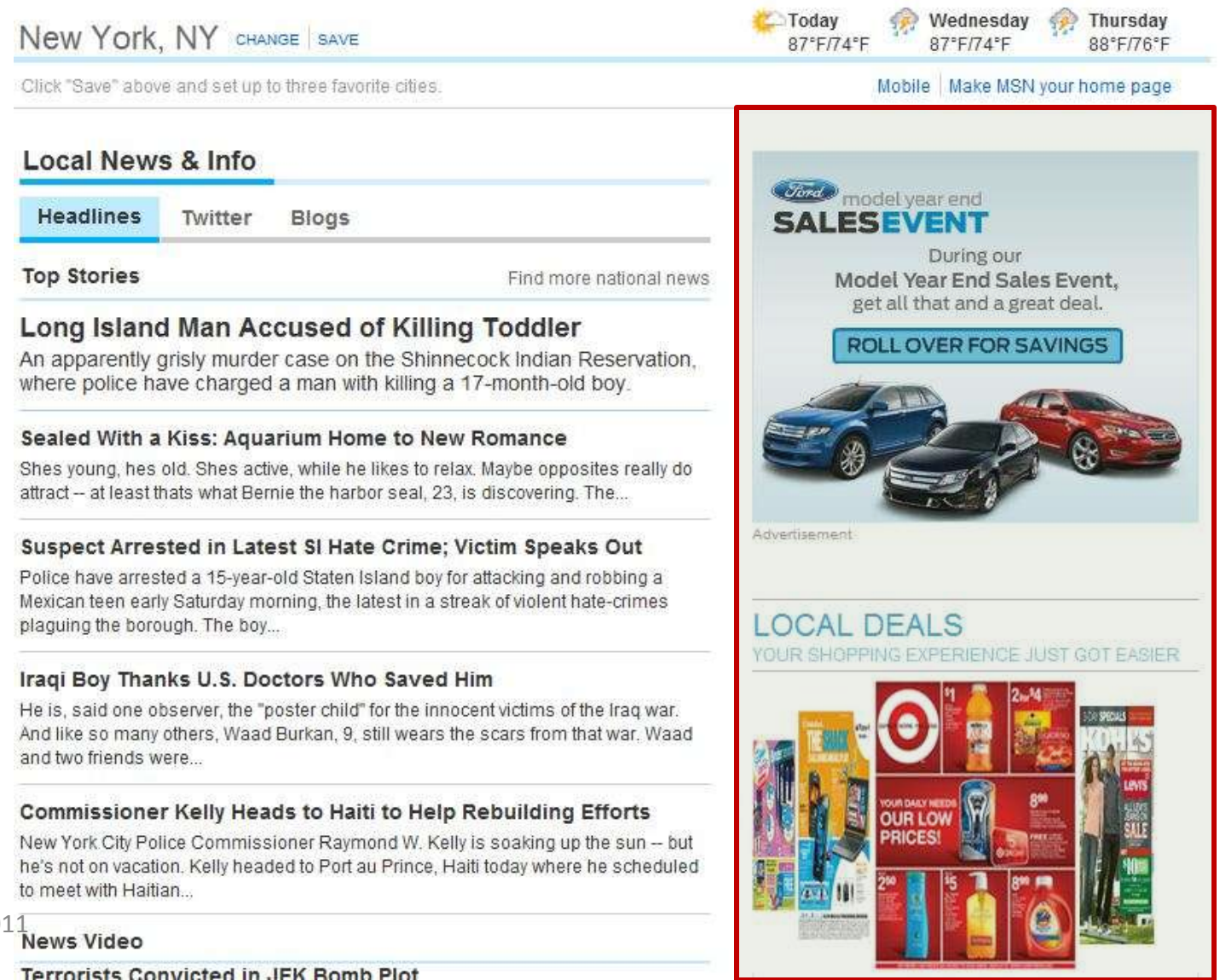
Types of Online Advertising

- Paid Search
  - Ads driven by search query issued by user
- Display Ads
  - Graphical ads primarily driven by behavioral targeting
- Contextual Ads
  - Ads driven by content of a page
- Mobile Ads
  - Mix of above types with focus on user profile and location
- Gaming Ads
  - Currently not focused deeply on matching

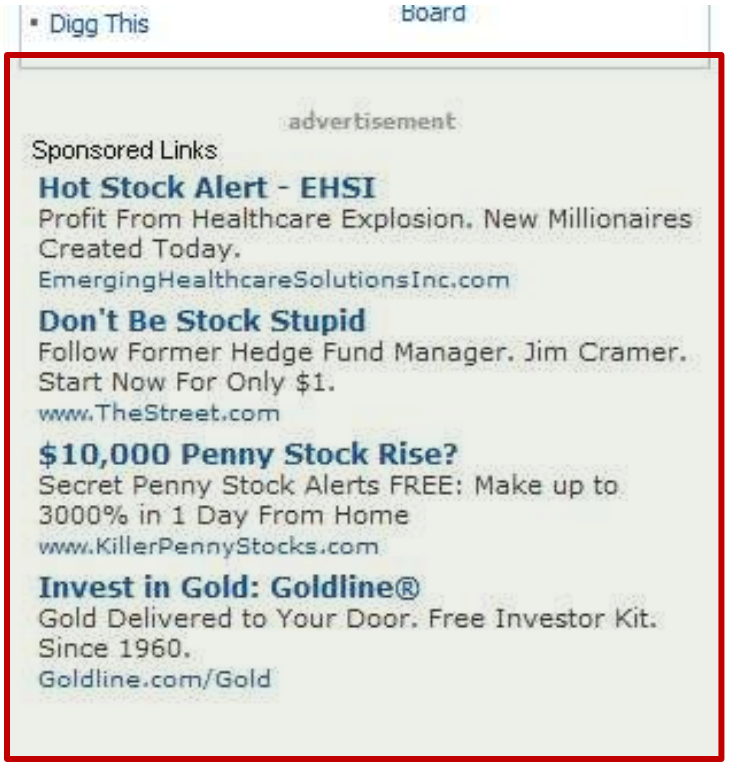
Paid Search


Web Images Videos Shopping News Maps More | MSN Hotmail
bing
Web
new york hotels
Web Visual Search Local Listings News Hotels Wikipedia
RELATED SEARCHES
Manhattan Hotels
Cheap Hotels New York
W Hotel New York
Roosevelt Hotel New York
New York City Hotel Guide
New York Cheap Hotels Directory
Waldorf Astoria
Sheraton New York Hotel Towers
SEARCH HISTORY
Search more to see your history
See all
Clear all · Turn off
ALL RESULTS
1-10 of 126,000,000 results · Advanced
540 New York Hotels · www.Expedia.com/NewYorkCity
Book your Hotels in New York on Expedia® - The #1 Site in Travel.
Marriott New York Hotels · Marriott.com/NewYork
Book Marriott's Best Rate Guarantee & Earn Rewards at Official Site.
250 Hotels in New York · Booking.com/Hotels-New-York-NY
Hotels online in New York. Book online now. Pay at the hotel.
Hotels in New York City · www.CandlewoodSuites.com
Pet Friendly Hotel In Times Square Unbeatable Location & Spacious Room
Find hotels in New York City (Manhattan)
Hotel Gansevoort
The Peninsula New York
The Plaza Hotel
The Benjamin Hotel New York
See Them All
New York Hotels - The Official New York Hotels .com Website
Book hotels at the OFFICIAL New York Hotels .com Website. With the help of thousands of user reviews, BEST PRICE guarantee and SECURE online reservations.
www.newyorkhotels.com · Cached page
Sponsored sites
Sponsored sites
New York City Hotel
A Botique Hotel With Personality. Unique Features - Weekends From $239
HotelRogerWilliams.com
Discount New York Hotels
Book online or call 1 800 522 9991 and save $$
www.hotelconxions.com
Over 540 New York Hotels
Good availability and great rates. Save up to 50% on your reservation!
www.Hotels.com
New York Hotels
Free Internet & Hot Breakfast Bar. Book Lowest Rates Direct!
www.HolidayInnExpress.com
New York Hyatt Hotels
Book Online for Hyatt's Best Hotel Rates on the Internet, Guaranteed.
www.NewYork.Hyatt.com
See your message here

Display Ads


New York, NY CHANGE | SAVE
Today 87°F/74°F
Wednesday 87°F/74°F
Thursday 88°F/76°F
Click "Save" above and set up to three favorite cities.
Mobile | Make MSN your home page
Local News & Info
Headlines
Twitter
Blogs
Top Stories
Find more national news
Long Island Man Accused of Killing Toddler
An apparently grisly murder case on the Shinnecock Indian Reservation, where police have charged a man with killing a 17-month-old boy.
Sealed With a Kiss: Aquarium Home to New Romance
Shes young, hes old. Shes active, while he likes to relax. Maybe opposites really do attract -- at least thats what Bernie the harbor seal, 23, is discovering. The...
Suspect Arrested in Latest SI Hate Crime; Victim Speaks Out
Police have arrested a 15-year-old Staten Island boy for attacking and robbing a Mexican teen early Saturday morning, the latest in a streak of violent hate-crimes plaguing the borough. The boy...
Iraqi Boy Thanks U.S. Doctors Who Saved Him
He is, said one observer, the "poster child" for the innocent victims of the Iraq war. And like so many others, Waad Burkan, 9, still wears the scars from that war. Waad and two friends were...
Commissioner Kelly Heads to Haiti to Help Rebuilding Efforts
New York City Police Commissioner Raymond W. Kelly is soaking up the sun -- but he's not on vacation. Kelly headed to Port au Prince, Haiti today where he scheduled to meet with Haitian...
News Video
Terrorists Convicted in JFK Bomb Plot
Ford
model year end
SALESEVENT
During our
Model Year End Sales Event,
get all that and a great deal.
ROLL OVER FOR SAVINGS
Advertisement
LOCAL DEALS
YOUR SHOPPING EXPERIENCE JUST GOT EASIER
OUR LOW PRICES!

Contextual Ads



Mobile Ads


Internet Ex
5:44
http://mobile.live.com/
Windows Live
FedEx
Track your packages at mobile.fedex.com
Hotmail
Messenger
Search (web, local, etc.)
Spaces
MSN (news, sports, more)
Sign In
(C)2008 Microsoft.
Privacy Statement
Back
Menu


the Nissan LEAF
the new car
DOVE MEN+CARE
Dove
MEN +CARE
iAd

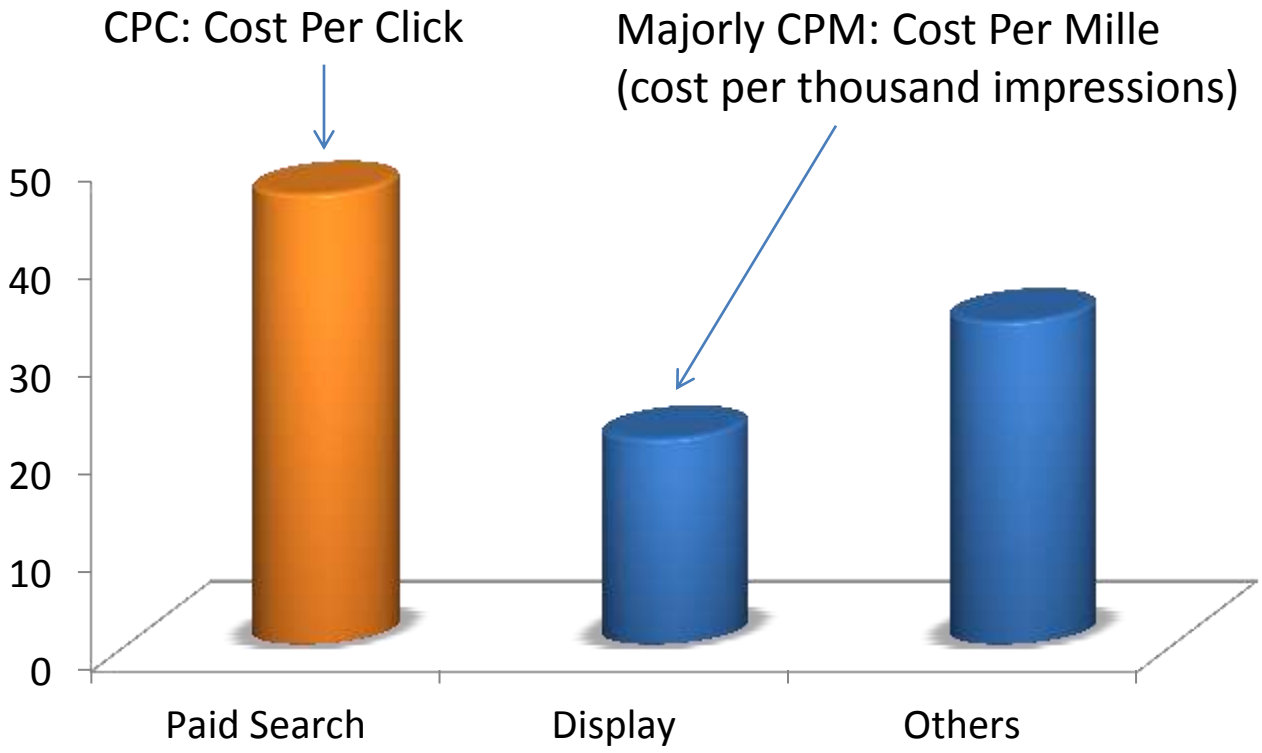
Gaming Ads


BOOSTUP
Ad Council
U.S. ARMY


NYY 0
ATL 0
1ST
88 MPH
0 OUTS
Schick
QUATTRO
TITANIUM


The Kia Soul
A new way to roll.
kiasoul.com
soul

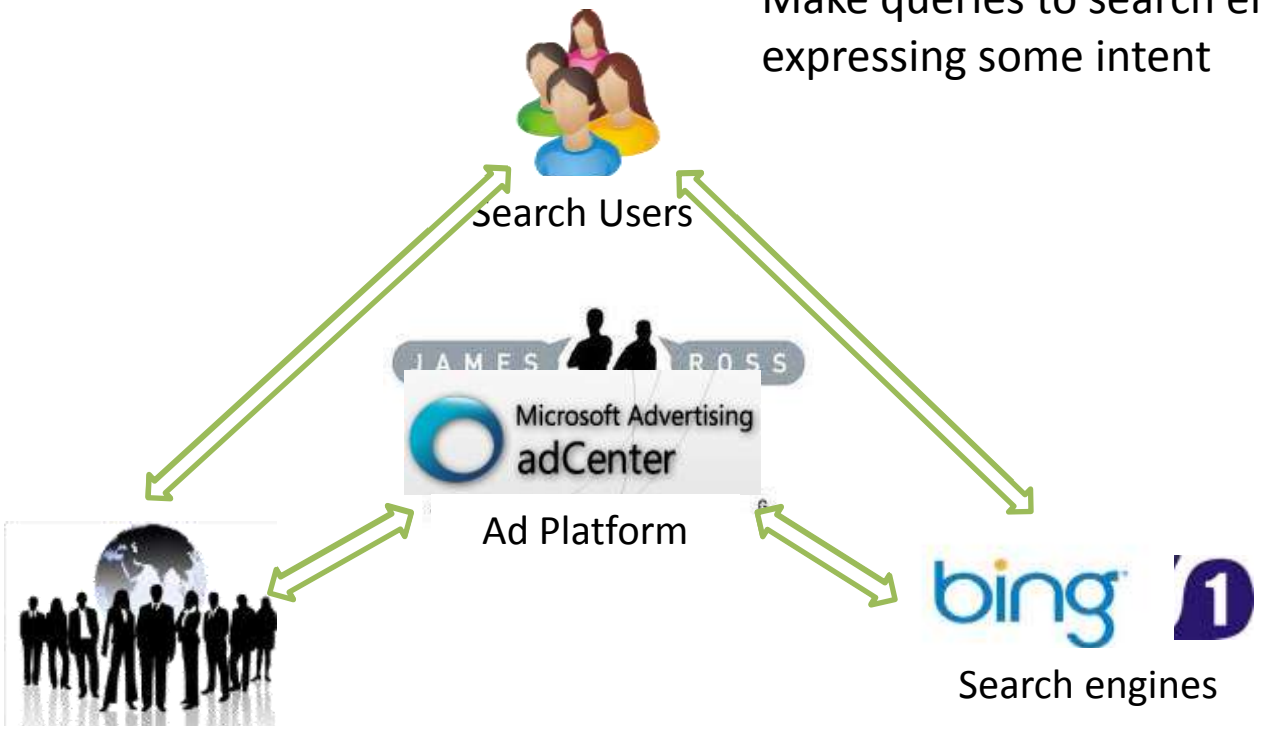
Types of Online Advertising



% of total revenue in Internet advertising

Paid Search: A Deep Dive

Eco-System in Paid Search



Overall Eco-system Utility

- Advertiser utility
  - Impressions, Clicks, CPC, Average position, Conversions
- Search engine utility
  - Revenue
- User utility
  - Relevance, usefulness

Facts about Paid Search

- Number of Advertisers: 1~2 Million
- Number of Ads: 500 Million ~ 1 Billion
- Number of Bid Keywords: ~1 Billion
- Search RPM: tens of USD
- Average CTR: 2~5%
- Ad Coverage in Search: ~50%
- Google's daily revenue in US: ~45 Million

User's View

Ad Copy

Title

Creative/ AdText

Display URL

Buy Microsoft SQL

We Have 6,700+ Programming Items.

Microsoft SQL on Sale!

www.NexTag.com/Programming-Software

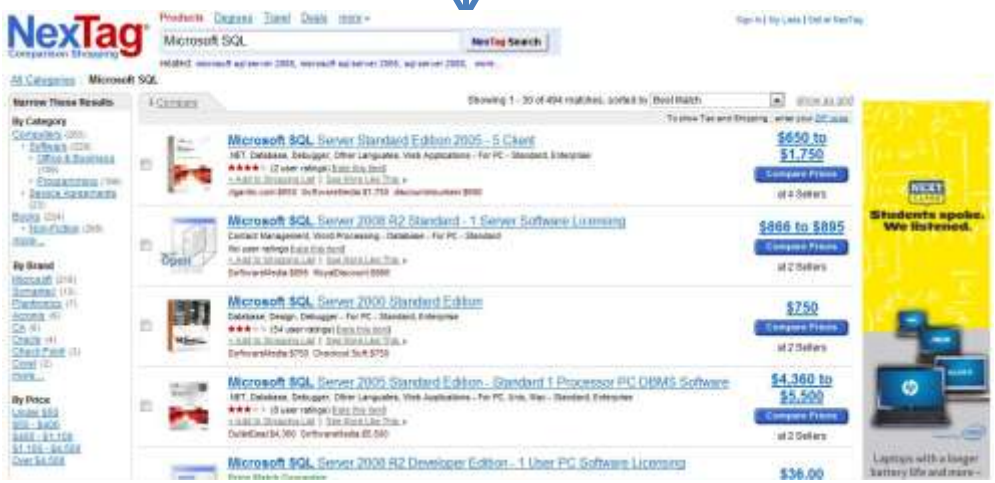
Bid Phrase: Microsoft SQL

Bid: $1.2

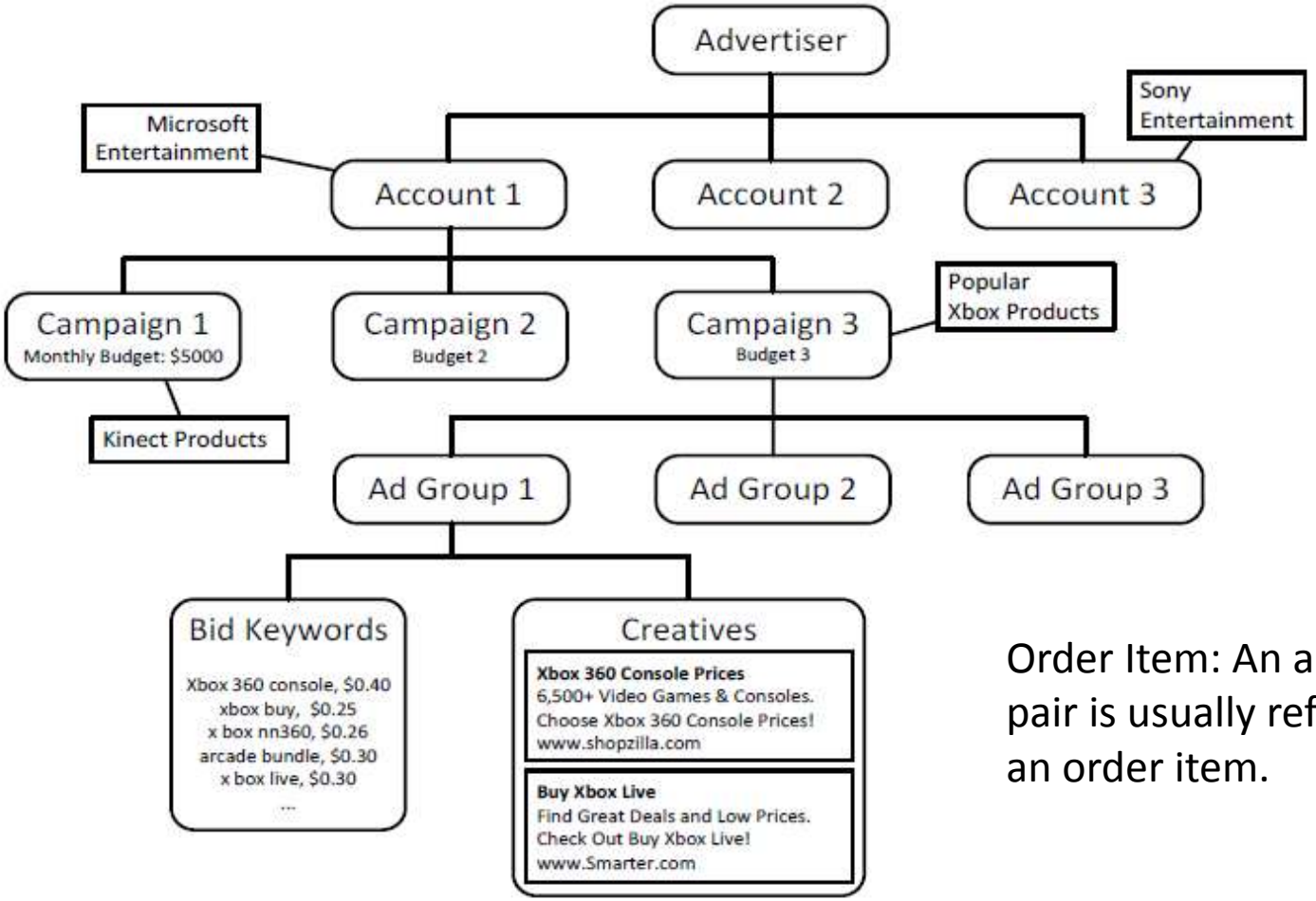
http://www.nextag.com/Microsoft-SQL/products-html?nxtg=4fdc0a28050c-9BCACF0210B560E9

Landing URL

Landing Page



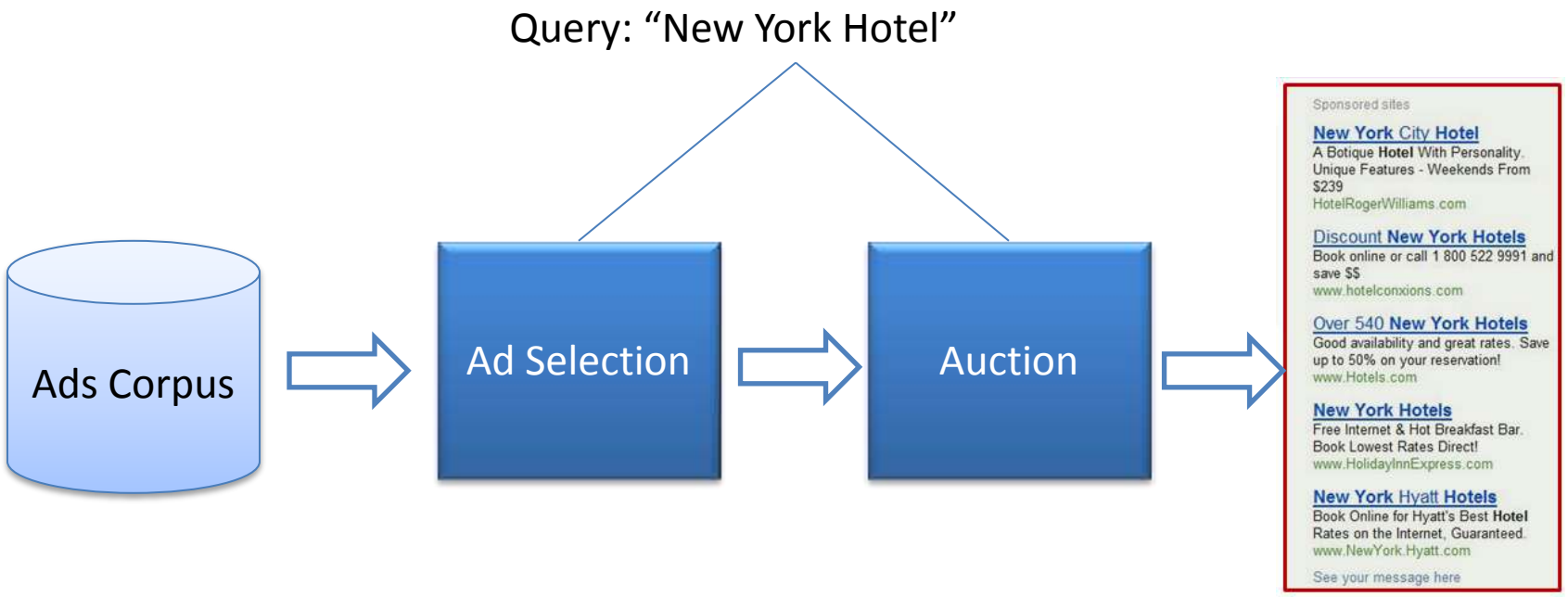
Advertiser's Campaign



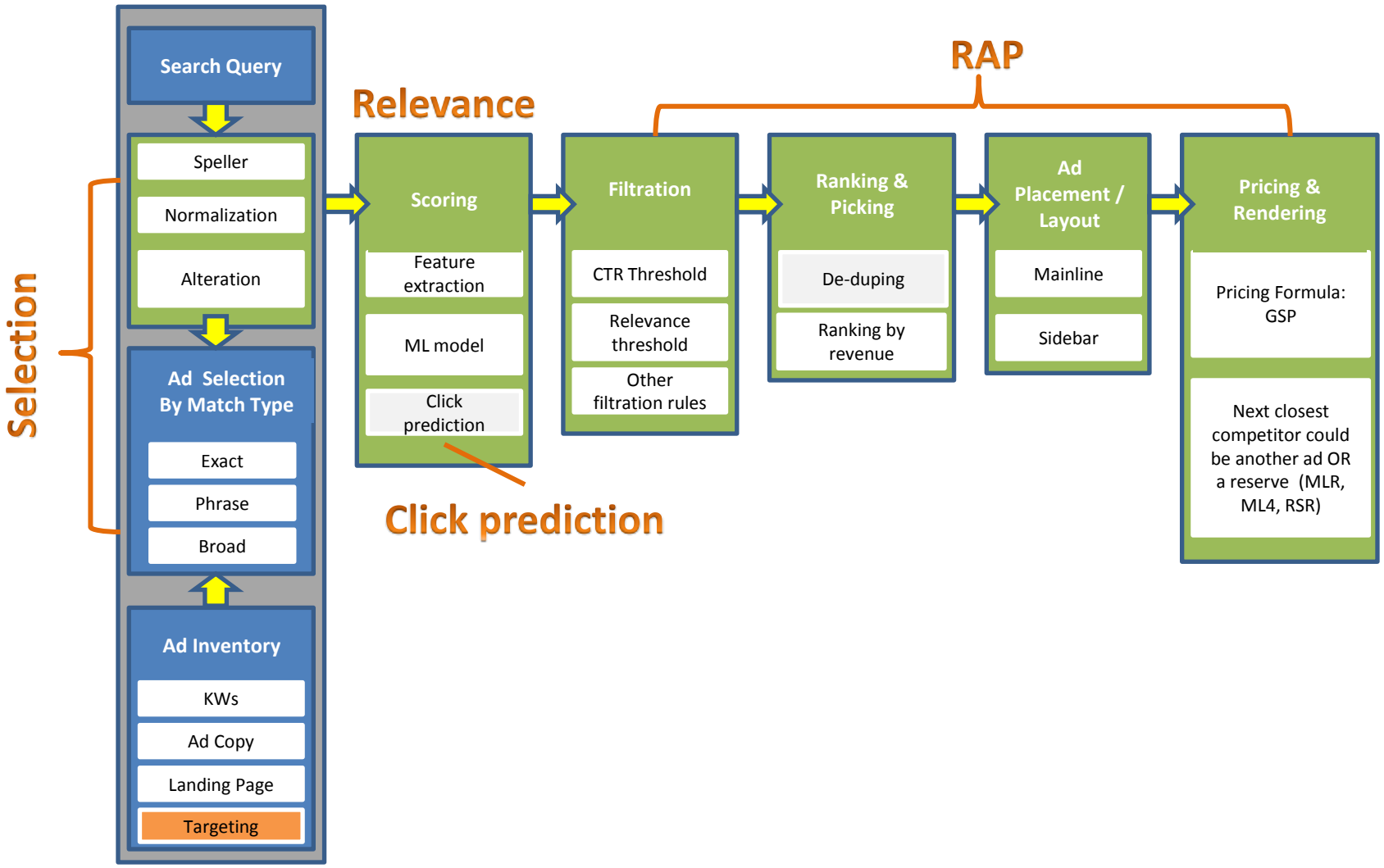
Order Item: An ad-keyword pair is usually referred to as an order item.

Ad Platform's Mechanism

- High-level pipeline



The Architecture



More Details

- Selection
- Relevance
- Click prediction
- RAP
- Evaluation

Selection

- The goal is to find those candidate ads that are potentially related to the query.

Match Type	Definition
Broad	Broad match includes words that are closely related to your keywords. For example, a search query for *red carnation* might result in your ad being displayed, because adCenter automatically identifies carnation as a type of flower. Use broad match to expose your ads to a wider audience.
Phrase	triggers the display of your ad if the word or words in your keyword appear in a customer's search query—even if other words are present in the typed query. Your keyword *red flower* would match searches for *big red flower* and *red flower*, but not *yellow flower* or *flower red*.
Exact	triggers the display of your ad only when the exact word or words in your keyword, in *exactly* the same order, appear in a customer's query. Your keyword *red flower* would *only* match searches for *red flower*, with no spelling variations. With exact match you might see fewer impressions (An ad that is served to and displayed on a user's browser.) but a higher click-through rate (The ratio of the number of times an ad is clicked to the number of times the ad is displayed.) , because your ad is shown to people who might be more interested in your product.

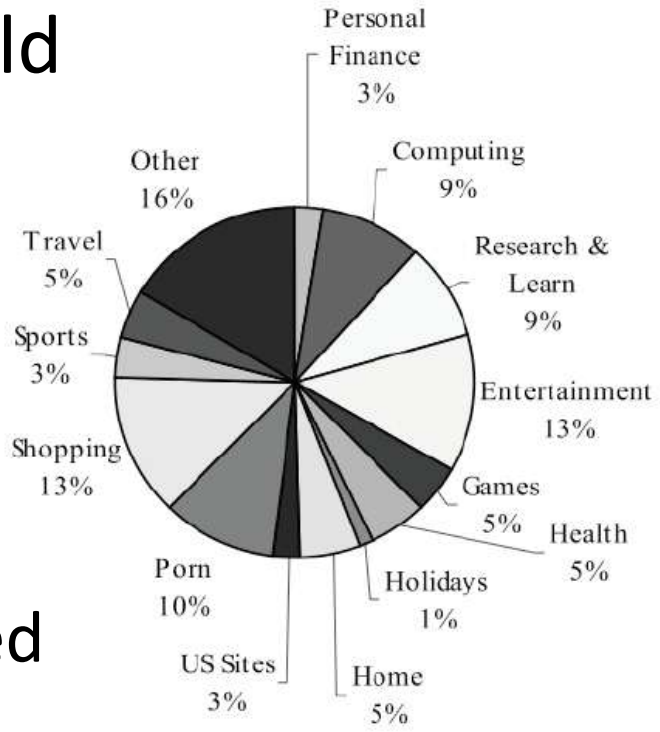
Broad Match

- Query rewriting / alteration: mostly a data-driven approach

	Query rewriting technique	Data source
1.	Generating Query Substitutions: Jones et al, in Proc of WWW 2006	query logs (query sessions)
	Using the Wisdom of the Crowds for Keyword Generation: Fuxman et al., In proc of WWW 2004	co-cliks on web search results
2.	Simrank++: Query Rewriting through Link Analysis of the Click Graph: Atoanellis et al., In proc of VLDB 2008	co-clicks on ads
3.	Learning Query Substitutions for Online Advertising: Broder et al. in Proc of ACM SIGIR 2008	query-to-ad similarity
4.	Online Expansion of Rare Queries for Sponsored Search: Broder et al, In Proc. of WWW 2009	query-to-query similarity
5.	Query Word Deletion Prediction: Jones at al., in Proc of ACM SIGIR 2003	query logs

Challenge: Select or Not?

- Showing ads is not always appropriate
- For some queries, we should even not select any ads.
  - To Swing or not to Swing: Learning when (not) to Advertise, CIKM 2008.
  - Estimating Advertisability of Tail Queries for Sponsored Search, SIGIR 2010.



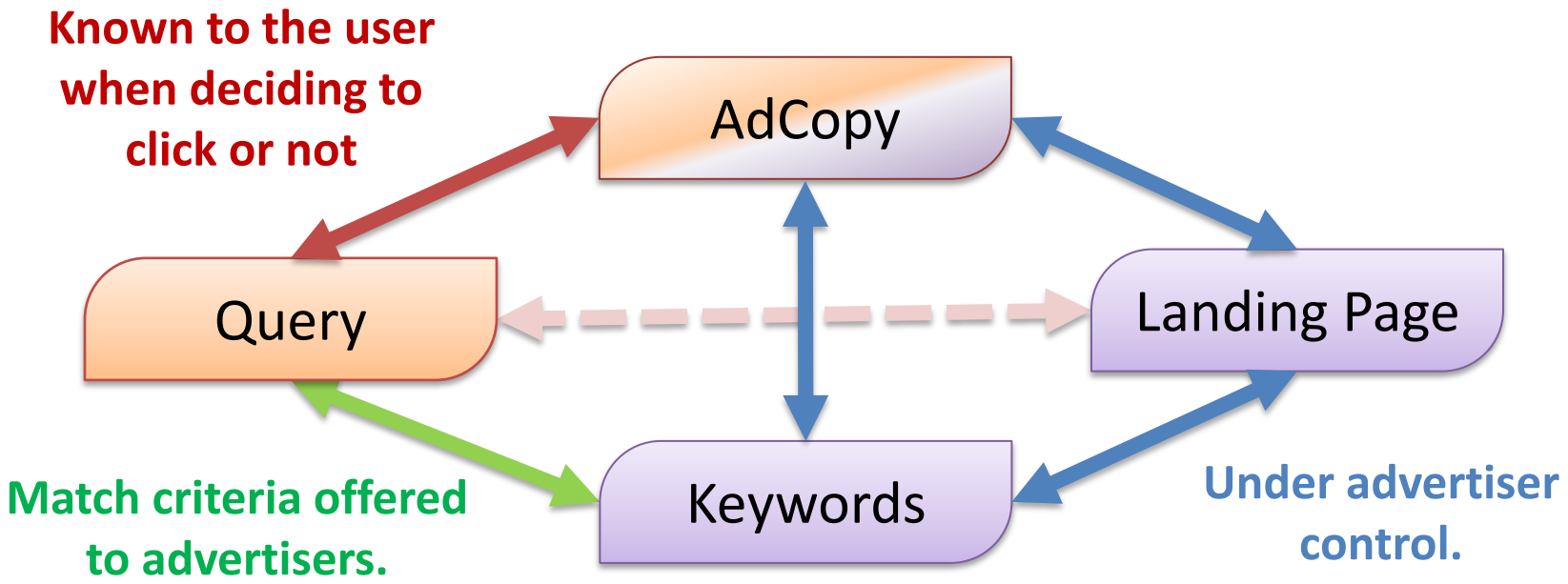
Challenges: Recall of Selection

- Recall is very important in the first step of ad retrieval due to economic considerations
- To guarantee recall, broad match is preferred
  - However, broad match does not work very well for tail queries, since usually tail queries do not have many log / session data.
  - For tail queries, semantic / syntactic analysis may work better.

Relevance

- **Relevance**: is a proxy for total utility:
  - Users – better experience
  - Advertisers – better (more qualified) traffic but possible volume reduction
  - SE gets revenue gain through more clicks but possible revenue loss through lower coverage

Relevance


Known to the user when deciding to click or not
AdCopy
Query
Landing Page
Keywords
Match criteria offered to advertisers.
Under advertiser control.

Ad Copy Relevance: Pre Click User Satisfaction

how to lose a guy in 10 days

Web Videos Images

ALL RESULTS 1-11 of 8,620,000 results · Advanced

Sponsored sites

Cheap Movie DVDs
HotMovieSale.com · **How To Lose A Guy In 10 Days** DVD. Yours For Only $2.29 + Free S&H!

How to Lose a Guy in 10 Days
www.howto**loseaguy**movie.com · Cached page · Mark as spam

How to Lose a Guy in 10 Days (2003) - IMDb
User rating: 6.1/10 · Comedy/Romance · 116 min
Andie Anderson covers the "How To" beat for "Composure" magazine and is assigned to write an article on "**How to Lose a Guy in 10 days**." They meet in a bar shortly after the bet is ...
www.imdb.com/title/tt0251127 · Cached page · Mark as spam

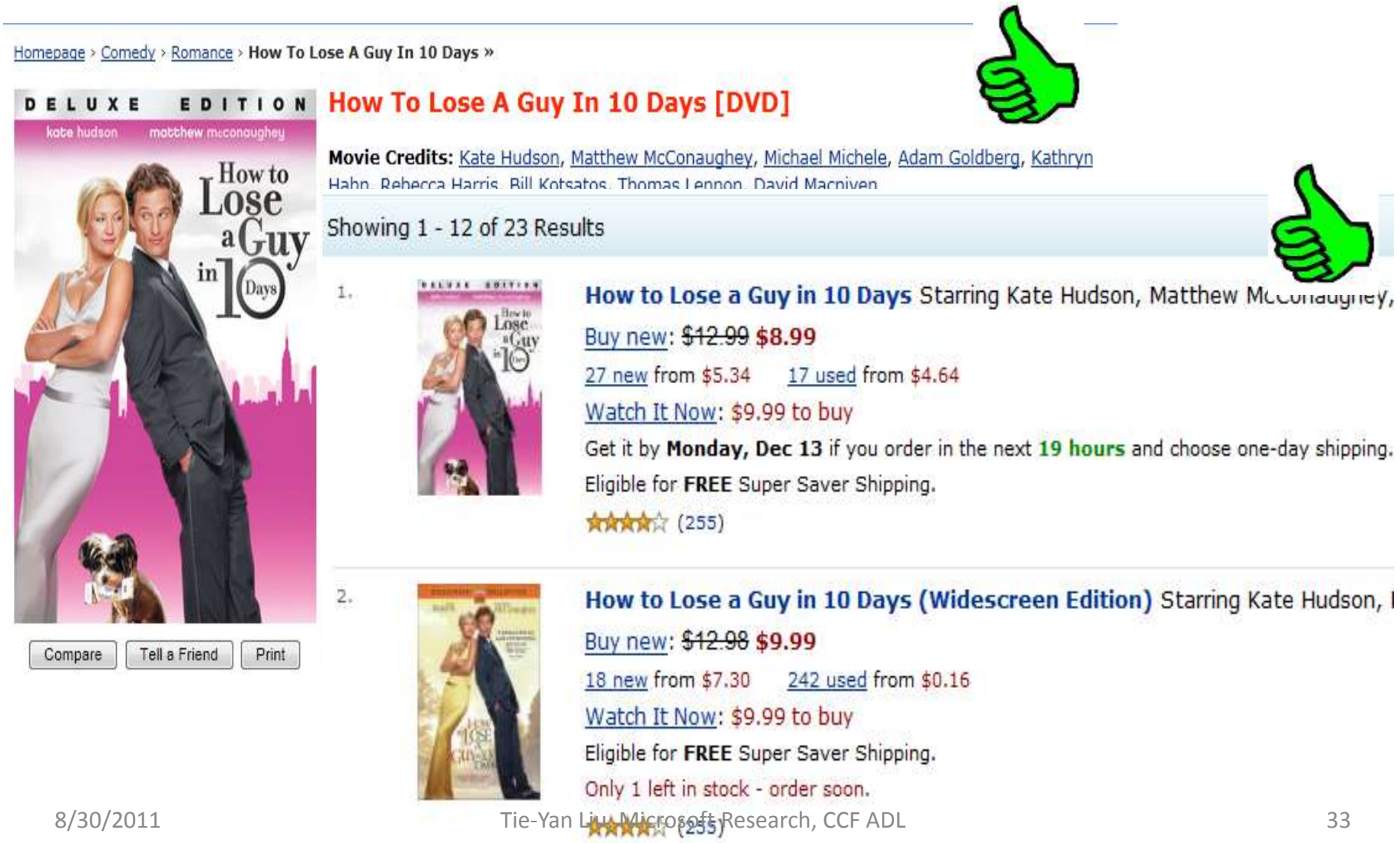
Sponsored sites

Jack Nicholson New Movie
With Reese Witherspoon, Paul Rudd & Owen Wilson - "**How** Do You Know."
HowDoYouKnow-Movie.com

Lose Guy 10 Days
Save up to 35% on top movies. Qualified orders over $25 ship free
www.Amazon.com/dvd

See your message here

Ad LP Relevance: Post Click User Satisfaction



Computing Relevance

- Machine learning approach
  - Extract features from query and ad copy/landing pages
  - Use a ML model (e.g., logistic regression) to learn from human relevance judgment
- Information retrieval approach
  - Index all ads copies and their landing pages
  - Use search system to retrieve ads

Challenges: Ad Copy Relevance

- Short text
- Term frequency cannot differentiate ads any more
- No anchor (which is one of the most useful signals in search)
- Both visible and invisible parts

Challenges: LP Relevance

- Classify landing page types for all the ads for 200 queries from the 2005 KDD Cup labeled query set.
  - **I. Category (37.5%):** Landing page **c**aptures the broad category of the query
  - **II. Search Transfer (26%):** Land on dynamically generated search results (same q) on the advertiser's web page
    - Product List – search within advertiser's web site
    - Search Aggregation – search over other web sites
  - **III. Home page (25%)**: Land on advertiser's home page
  - **IV. Other (11.5%):** Land on promotions and forms

[H. Becker, AB, E. Gabrilovich, VJ, B. Pang, SIGIR 2009]

Challenges: Is Relevance Crucial?

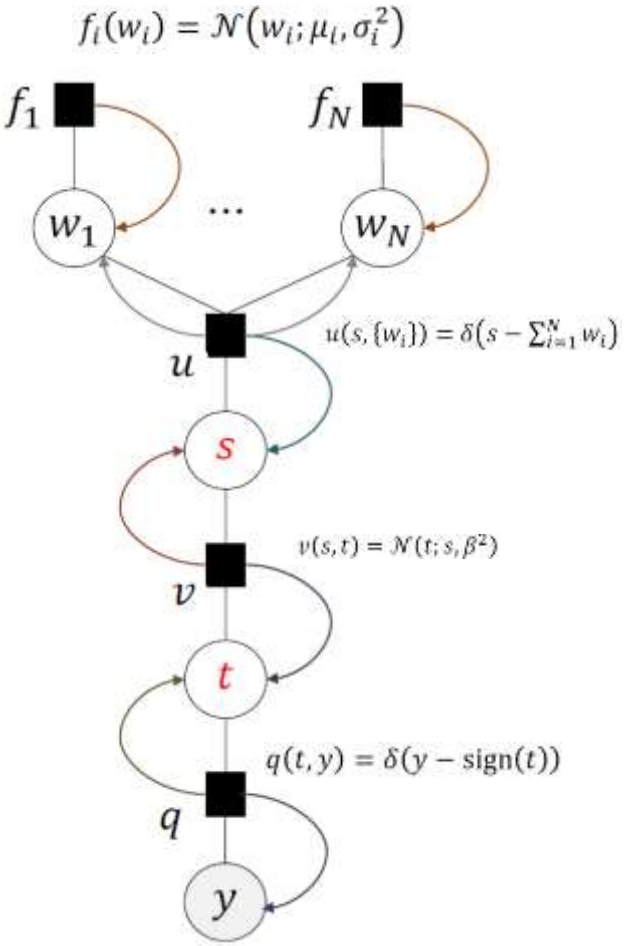
- Sometimes, the user satisfaction on an ad is not just determined by its relevance.
  - Informativeness
  - Attractiveness
  - Real discount
  - Spam…

Click Prediction

- Since advertisers are charged per click, an accurate prediction of click will be very important for ranking and pricing.
- Representative works
  - [ICML10] Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine
  - [ADKDD09] Data-driven text features for sponsored search click prediction
  - [ADKDD10] Missing Click History in Sponsored Search A Generative Modeling Solution
  - [WWW07] Predicting clicks estimating the click through rate for new ads
  - … …

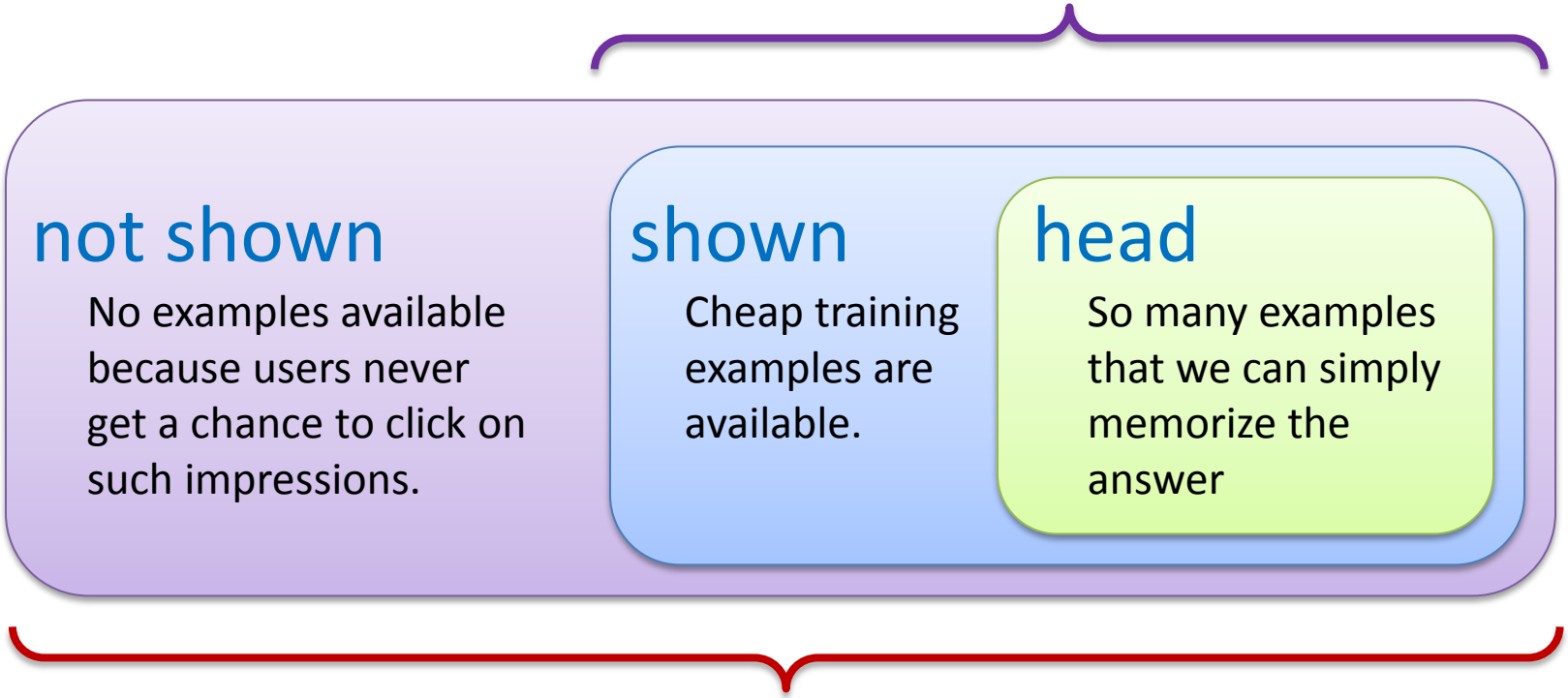
Examples

- adPredictor in Bing
  - Bayesian approach
  - A Sparse Linear Probit Regression
- pClick model in Yahoo!
  - A logistic regression based model
  - Leveraging the hierarchical structure of advertiser data to handle sparseness
  - Use GMM and EM to handle missing data



Challenges: Exploration

We only have **training data** for building models valid for these ad impressions.



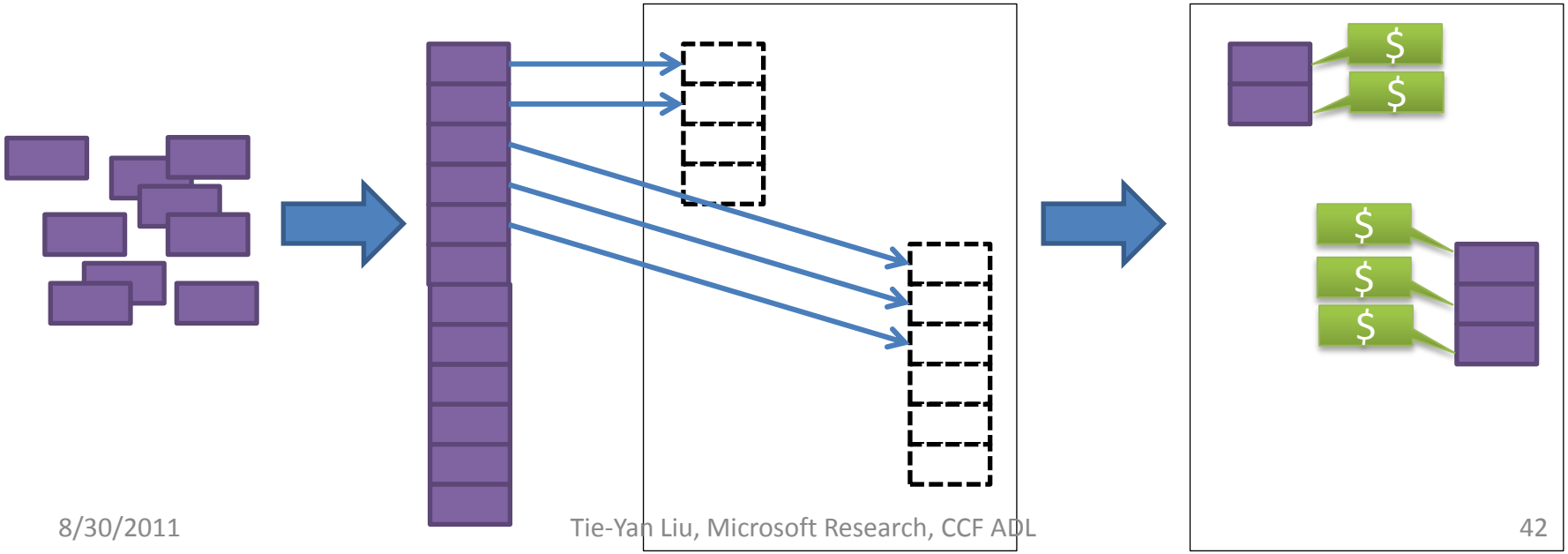
But we need a model that works for all potential impressions (because we need to know which ones should not be shown.)

Challenges: Second-order Effect

- Second-order effect: Result -> Cause -> Result
  - The pClick model is learned from historical data, reflecting user's click behaviors with respect to the previous ranking algorithm
  - After a new model is learned, users will see different ads and their behaviors may change correspondingly.
  - The gradual change of user behaviors will make the offline evaluation on the learned pClick model not valid when put online.

The RAP Problem

Ranking	Allocation	Pricing
• Sort the candidate ads	• Say where on the page each ad goes	• Set CPC prices (cost per click)



RAP: Ranking, Allocation, Pricing

- **Ranking**: sort the candidate ads
  - RS = Bid x pClick
- **Allocation**: say where on the page each ad goes.
  - Mainline Reserve
  - Overall Reserve
- **Pricing**: Generalized Second Price

$$CPC_i = Bid_{i+1} \times \frac{pClick_{i+1}}{pClick_i}$$

RAP Mechanism

- Second price auction
  - All advertisers submit their bids privately
  - Ads are ranked according to their bid prices
  - The ad ranked on the first place will win the auction; the owner of the ad will pay a price slightly higher than the price of the second highest bid.
- Generalized second price auction
  - “Pay-per-click” pricing: buy placement but not pay for placement!
  - Ads are ranked in the order of CTR * bid
  - The winner pays the cost of maintaining his/her position, if clicked

Factors to Consider for a Mechanism

- Incentive compatibility (truthfulness)
- Equilibrium
- Social welfare
- Auctioneer revenue
- Approximation ratio

Incentive Compatibility

- A process is said to be *incentive compatible* if all of the players fare best when they truthfully reveal any private information asked for by the mechanism
- In particular, an auction mechanism is **truthful**, if the dominant strategy for every bidder is to truthfully bid their own value on the item.

- SP is truthful, FP and GSP are not truthful.

Equilibrium

- **Nash equilibrium** = choice of strategies in which each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing his/her own strategy unilaterally.

- For GSP, there exist many Nash equilibrium strategies.

Social Welfare

- $\sum_i \ U_i(x_i)$

- The utilities can be
  - Advertiser ROI (true value – cost)
  - Clicks /CTR
  - Impressions
  - Conversion rate

- In most previous work, auction mechanisms are designed to maximize social welfare.

Auctioneer Revenue

- How much does the auctioneer actually get from the auctions?
  - E.g., $\sum_i \ CTR(x_i)Prc(x_i)$
- In some previous work, mechanisms are designed to optimize auctioneer revenue.
- GSP is not revenue optimal.
  - We can increase RSP by finding a revenue-optimal mechanism.

Approximation Ratio

- The problem of finding the optimal allocation is usually NP-hard.
- An approximation algorithm with polynomial time complexity is desirable.

- Approximation ratio is related to the reduction of the objective function (social welfare and/or auctioneer revenue), introduced by the approximation algorithm.

Challenges: Simple vs Optimal

- We can design an optimal mechanism by considering many different requirements, however, the optimal mechanism might be complex.
- Advertisers prefer simple mechanism that they can understand and can play with.

Challenges: Assumptions

- Many researches on incentive compatibility and equilibrium assume
  - Advertisers are rational
  - Advertisers know their true values
  - Advertisers have good access to other advertisers' information.
- However, in reality, it is often not the case.
  - When these assumptions do not hold, the theoretical results might not be applicable to advertising practices.

Challenges: Beyond Traditional Auctions

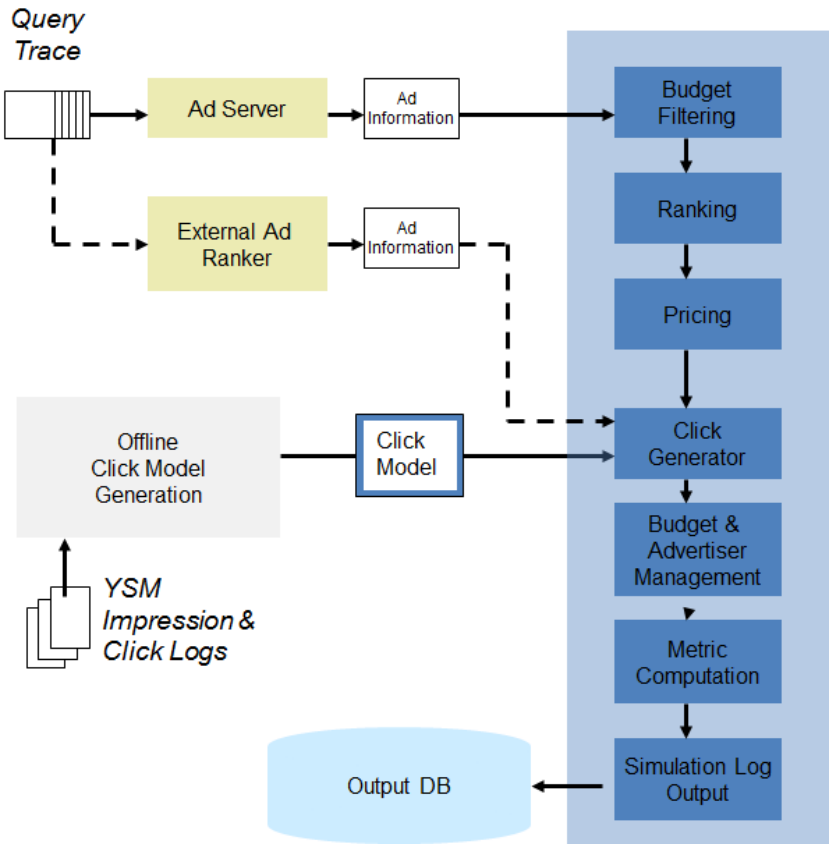
- Real paid search auctions are
  - Repeated auctions
  - Broad-match auctions
- Paid search auctions can also be
  - CPA auctions
  - Hybrid auctions (CPM + CPC + CPA)
  - References
    - Hybrid Keyword Search Auctions, WWW 2009
    - Dynamic Cost-Per-Action Mechanisms and Applications to Online Advertising, WWW 2008

Evaluation

- Online evaluation
  - Live test itself is risky
    - A failed configuration will decrease our revenue
    - May damage our brand impression
  - The fraction of live traffic is typically small to mitigate risk
    - The collected data/observations may be insufficient
    - The results may be not reliable
    - Cannot run many configurations simultaneously
    - Long time period is needed to accumulate data
  - Isolating all design factors in a production system is usually not possible
    - Difficult to draw conclusions for the investigated factor

Evaluation

- Offline evaluation
  - Entire traffic
  - Implement faster
  - Easy to isolate design factors
  - Create scenarios to model advertiser consequences



Cassini in Yahoo!

Evaluation

- Challenges of offline experiments
  - Efficiency
  - Evaluation measures
  - Dynamics in users and advertisers
  - Second-order effects

Summary

Online Advertising is Promising!

- It is about a business model, but not just a component technology!
- It is cross-discipline, with a lot of unexploited directions!
  - Information retrieval
  - Machine learning
  - Game theory
  - Economics
  - … …

Online Advertising is Hard!

- It is cross-discipline, and therefore requires a wide range of knowledge and skills
- It can be highly theoretical: needs deep understanding of game theory and computational theory
- It can be highly practical: needs a lot of data support and needs to be verified by real systems.

Let's Do It!

- As researchers, we like to work on promising directions.
- As researchers, we like to solve challenging problems.

Thanks!

tyliu@microsoft.com

http://research.microsoft.com/users/tyliu/

http://weibo.com/tieyanliu