

Sponsored Search Tutorial

Kartik Hosanagar
The Wharton School

Michael Schwarz
Yahoo Inc

Outline

- ◆ “Science will replace much of the art of marketing” – Eric Schmidt
- ◆ Introduction
 - Online Advertising and SEM basics
- ◆ Search Engine Perspective in SEM
 - Mechanism Design and Auction Rules
 - Challenges
 - Click Fraud
- ◆ Advertiser Perspective
 - Keyphrase Generation
 - Bidding
- ◆ Sponsored Search with Context
- ◆ Empirical Issues in Sponsored Search
- ◆ Conclusions

Introduction

Traditional Advertising

- ◆ Print, TV, Radio, Direct Mail
- ◆ Ad agencies buy media on behalf of firms
 - Dominated by large holding companies (Omnicom, WPP, InterPublic, Publicis Groupe)
 - Individual agencies like JWT, Ogilvy & Mather, McCann Erickson manage branding
 - Low barriers to entry (talent and sales)
 - » But some economies of scale
- ◆ Local Advertising (YPs)
 - Publisher centric

Traditional Advertising

- ◆ Pricing
 - Traditionally fixed commissions
 - Later negotiated commissions
 - Recently, labor-based + performance-based
 - » Pushed by advertisers: aligns incentives
 - » Reflects growth in promotions, event marketing, product placement

Advertising spend

Media	2006
Newspapers	47
Direct Mail	56
Broadcast TV	45
Radio	20
Cable TV	22
Magazine	16
Yellow Pages	15
Internet	17
TOTAL	238

Online advertising

- ◆ Banner ads (DoubleClick)
 - Standardized ad shapes with images
 - Loosely related to content
- ◆ Search linked ads (Google Adwords)
 - Related to search terms
- ◆ Context linked ads (Google AdSense)
 - Related to content on page
 - Closely tied to Google's existing competence on figuring out what a page is really about

Search Engine Marketing (SEM)

The screenshot shows a Mozilla browser window displaying Google search results for the query "search engines". The browser's address bar shows "http://www.google.com/ser...". The search results page includes a "Sponsored Link" for "All Major Search Engines" and several organic search results. Annotations with arrows point to the "Sponsored ad" (the top result) and "Search results" (the organic results below).

Sponsored ad

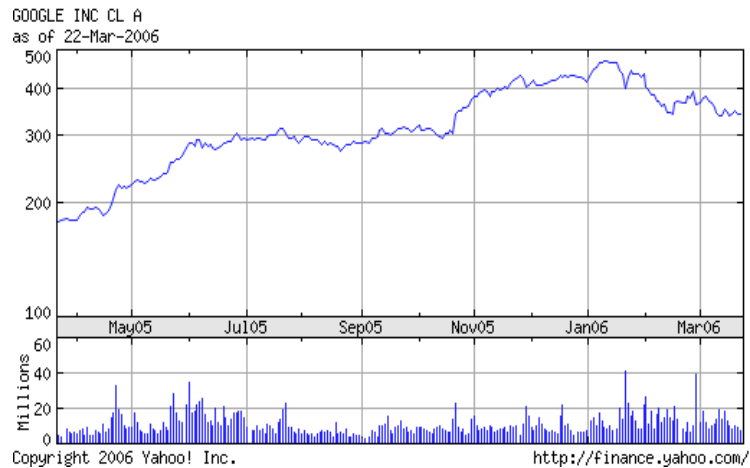
Search results

Sponsored ads

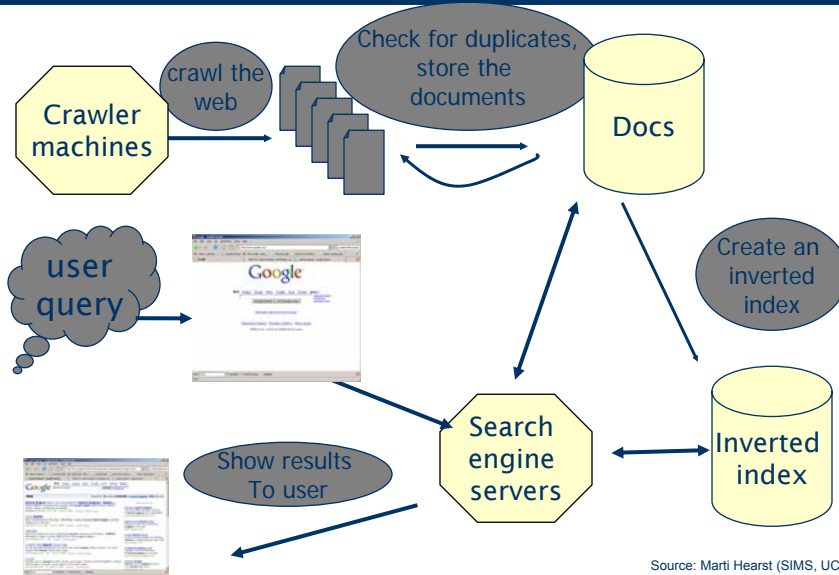
Search Engine Marketing

- ◆ High relevance (based on user context)
- ◆ Occurs close to decision/buying time
- ◆ Highly measurable
- ◆ Results in very good performance in cost-per-acquisition (adv expenditure per sale)
- ◆ Rapid growth but still small part of advertising market
 - Total US adv spend ~ \$220-240 billion, growth ~ 1-2% per year
 - Online advertising: ~ \$17 billion, growth ~ 35% last year

Google Valuation

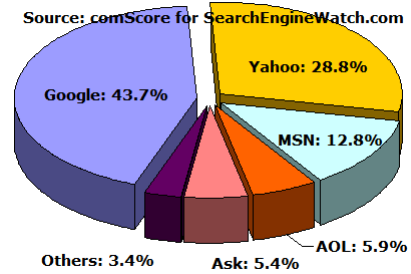
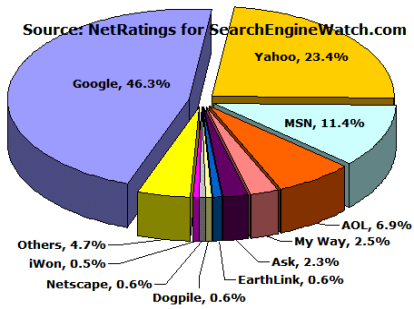


Information Retrieval (IR)



Source: Marti Hearst (SIMS, UC Berkeley)

Share of Searches



- ◆ Organic Traffic is a key to driving SEM revenues

Source: Search Engine Watch (2006)

How does SEM work?

- ◆ Search engines
 - run keyword auctions to sell available inventory of ad positions
- ◆ Advertisers
 - submit bids which indicate their willingness-to-pay per click
 - » for example, bid of \$2.10 per click for the keyword “laptop”
- ◆ The search engine orders the ads in descending order
 - Bid is a key determinant of ad position
 - Other factors such as CTR are also factored in
 - » More on the mechanism later

Search Engine Perspective

The Advertiser Perspective

Advertiser Perspective

- ◆ Choose keywords
- ◆ Set budget
 - Budgets are commonly observed
 - Proxies for inventory constraints, short-term capital constraints, etc
 - Sometimes, to protect against click-fraud
- ◆ Determine bid (maxCPC) for each keyword
 - Auctions not incentive compatible
 - Budget constraints may restrict bidding strategies
- ◆ Role of SEM firms

Keyphrase Generation

Current Methods

- ◆ Query log and Advertiser log mining
 - Mine query logs and logs of advertiser searches
 - » Google and Yahoo
 - Mine other keyphrases advertisers have bid on
 - » Yahoo
- ◆ Proximity search
- ◆ Meta-tag spidering
 - Extract meta-tags from top web pages related to seed keyword
 - Word tracker
- ◆ Terms Net (Joshi and Motwani 2006)
 - Based on semantic similarity between keyphrases

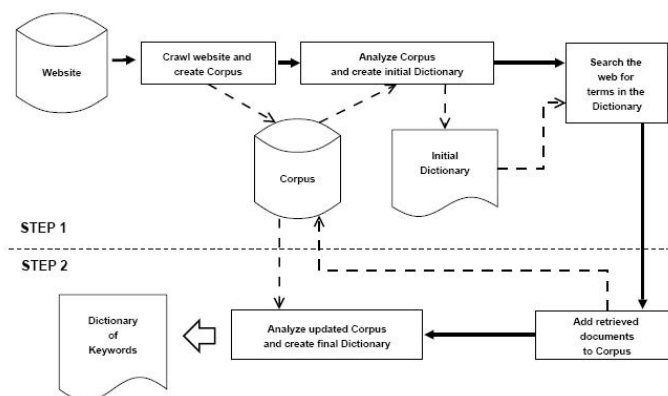
Semantic Techniques for Keyphrase Generation

An automated system for relevant keyword generation.

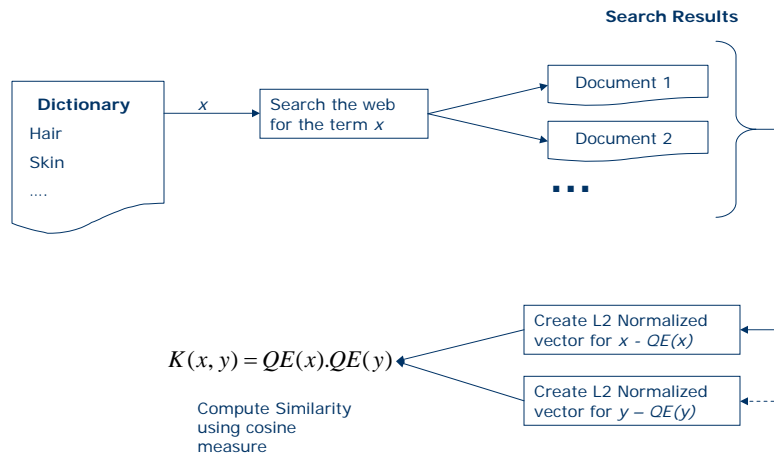
Three Steps:

- ◆ Dictionary Creation
- ◆ Semantic Similarity
- ◆ Keyword Suggestion

Dictionary Creation



Semantic Similarity



Keyword Generation

- ◆ Construct similarity graph using the association matrix.
 - Edge weight inversely related to similarity score
- ◆ Traverse the graph using the watershed algorithm.
- ◆ Assumption – words with lower frequency are cheaper.

Examples

skin

skincare	body	aromatherapy
facial	scalp skin	hair removal
treatment	skin smooth	massage
face	scalp skin	healthcare
care	skin smooth	microdermabrasion
occitane	acid peels	natural
exfoliator	free cosmetics	permanent makeup
dermal	makeup	aging
wrinkles	conditioner	shaving
dermal	wrinkles	treatment glycolic

Metrics

- ◆ Precision: Fraction of keyphrase recommendations that are deemed relevant
- ◆ Recall: Fraction of relevant keyphrases identified by system
 - Universe of relevant keyphrases is unknown
 - Relative measures can be obtained by implementing multiple keyphrase generation techniques
 - » Treat union of relevant recommendations as the universe
- ◆ Other economic criteria observed empirically
 - Changes in revenues, profits, average CPC, etc

Budget Constrained Bidding

A Toy Problem

- ◆ The advertiser has a daily budget of \$1000
- ◆ She is considering bidding on four keywords
 - Laptop
 - Refurbished Laptop
 - HP Laptop
 - Sony Laptop
- ◆ Assume the bids submitted by competitors are known to the advertiser
- ◆ How should the advertiser bid to get the maximum number of clicks (visitors from the search engine ads) each day?
- ◆ In real life 1,000s of keywords and millions of \$ per month...

Bid Landscape(\$ per click)

- ◆ Advertiser *i*'s view of other advertisers' bids for the same keywords
 - found via query of Yahoo or Google's advertising bid database

Slot	Laptop	Refurbished Laptop	HP Laptop	Sony Laptop
1	6.73	1.29	0.59	0.93
2	6.72	0.95	0.58	0.85
3	2.00	0.80	0.56	0.84
4	1.70	0.79	0.55	0.70
5	1.56	0.78	0.51	0.58
6	1.42	0.77	0.42	0.46
7	1.13	0.57	0.40	0.43
8	0.95	0.54	0.40	0.39

Expected number of clicks for each keyword and slot

- ◆ Advertiser *i*'s view of the benefit of obtaining a given slot for a keyword
 - found via query of Yahoo or Google's advertising bid database
 - and through advertiser's own historical clickthrough data

Slot	Laptop	Refurbished Laptop	HP Laptop	Sony Laptop
1	1539	370	45	38
2	1099	264	32	27
3	785	189	23	19
4	561	135	16	14
5	401	96	12	10
6	286	69	8	7
7	204	49	6	5
8	146	35	4	4

Tradeoffs to make when deciding what to bid?

Slot	Laptop		Refurbished Laptop		HP Laptop		Sony Laptop	
	Bid	Clicks	Bid	Clicks	Bid	Clicks	Bid	Clicks
1	6.73	1539	1.29	370	0.59	45	0.93	38
2	6.72	1099	0.95	264	0.58	32	0.85	27
3	2.00	785	0.80	189	0.56	23	0.84	19
4	1.70	561	0.79	135	0.55	16	0.70	14
5	1.56	401	0.78	96	0.51	12	0.58	10
6	1.42	286	0.77	69	0.42	8	0.46	7
7	1.13	204	0.57	49	0.40	6	0.43	5
8	0.95	146	0.54	35	0.40	4	0.39	4

Problem formulation

- ◆ Decision variables
 - how much to bid on each of four keywords
- ◆ Objective function
 - maximize total number of clicks
- ◆ Constraints
 - budget – maximum \$1,000 spent
 - assignment constraints
 - » at most one bid for each key word

Decision variables and objective for the problem

◆ *Decision Variables*

$x_{ij} = 1$ if slot i is assigned to keyword j , 0 otherwise;

where $i = 1, 2, \dots, 8$ represents the eight slots
and $j = A, B, C, D$ represents the 4 keywords.

◆ *Objective Function*

$$\text{Max } 1539x_{1A} + 370x_{1B} + 45x_{1C} + 38x_{1D} + 1099x_{2A} + \dots + 4x_{8C} + 4x_{8D}$$

Constraints for the IP

◆ *Each keyword must be assigned to one, and only one slot:*

$$x_{1A} + x_{2A} + x_{3A} + \dots + x_{8A} \leq 1 \quad (\text{Keyword A})$$

$$x_{1B} + x_{2B} + x_{3B} + \dots + x_{8B} \leq 1 \quad (\text{Keyword B})$$

⋮

$$x_{1D} + x_{2D} + x_{3D} + \dots + x_{8D} \leq 1 \quad (\text{Keyword J})$$

◆ *Total Budget for the day = \$1000:*

$$x_{1A} \cdot b_{1A} + x_{1B} \cdot b_{1B} + \dots + x_{8D} \cdot b_{8D} \leq 1000$$

where b_{ij} is 1 cent more than the bid of the advertiser currently ranked i for keyword j

◆ *Integrality*

x_{ij} are all either 0 or 1

Issues with Formulation?

- ◆ Bid landscape typically unknown
- ◆ Clicks Vs Position unknown
- ◆ Keywords are not of same quality
 - Conversions matter
- ◆ IP not scalable

Approach 2: Knapsack

- ◆ Knapsack problem and greedy heuristic
- ◆ Adaptive algorithm for keyword selection (Paat R* & Williamson)
 - Static model: case where the CTR for the keywords is known
 - » Sort keywords in descending order of profit/cost.
 - » Choose keywords in descending order (of profit/cost) until the expected cost is close to the budget.
 - Adaptive algorithm works when CTR is not known apriori
 - » Use data gathered in the first few periods to estimate the CTR

- ◆ Issues: Does not account for slots

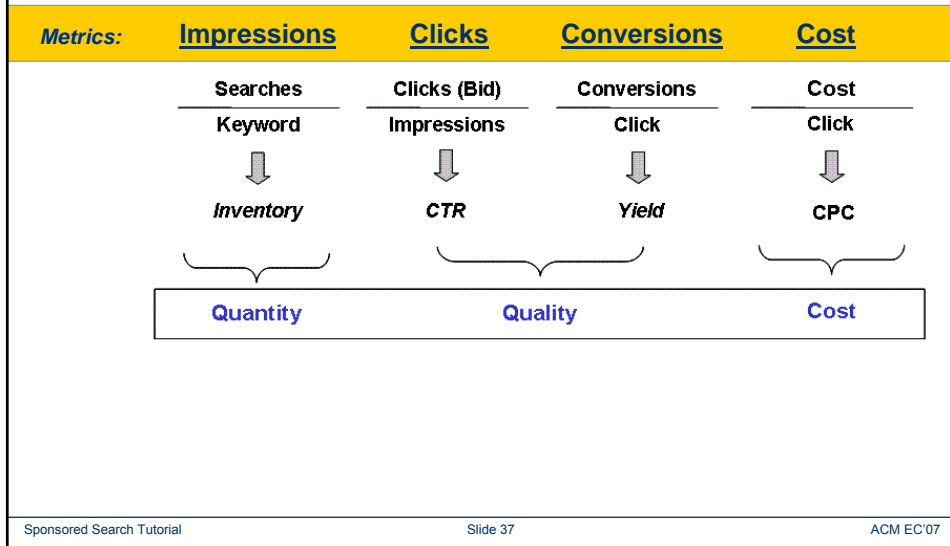
Other Techniques for Bidding

- ◆ Online knapsack Problem (Chakrabarty et al. 2007)
- ◆ Stochastic bidding (Muthukrishnan et al. 2007)
- ◆ Efficient Frontier (Holthausen and Assmus 1982)
- ◆ Convex Programming: Scalable and accounts for slots (Hosanagar and Stavrinides 2006)

Role of Machine Learning

- ◆ “Standard” Optimization techniques solve deterministic problems
- ◆ Even stochastic optimization techniques generally assume parameters can be estimated
- ◆ Parameters
 - Bid-Rank
 - Rank- Clicks
 - Clicks-Conversions
 - Bid-CPC

Role of Machine Learning



Real-world Complexities

- ◆ “Exact match”
 - Competing with other advertisers you do not know of
- ◆ “Broad Match”
 - Your rank changes based on the search query
- ◆ Data Integrity
 - Impressions < clicks
 - CPC > Bid
- ◆ Data Sparseness

Understanding the Role of SEM Firms

Advertiser

- ◆ Unable to easily generate business online
- ◆ Search auctions are too complex for many advertisers
- ◆ Bidding Blindly on Search Engines
- ◆ RoI hard to measure
- ◆ Publishers User-centric, not business-centric

- ◆ SEM firms use a variety of techniques to
 - Identify keywords to bid on
 - Bid for each keyword
 - Provide metrics

SEM Firms' Value Proposition

Aggregating the fragmented search channels

Online Users



Generating Keywords

Configure»

Keywords Created

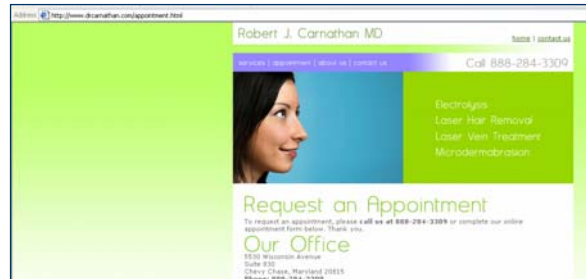
cholesterol
heart medication
blood pressure medication
aspirin
diabetes
diet
obesity
electric help in philadelphia

Determining Bids

Slot	Cholesterol	Blood Pressure	Lipid Reduction	Heart Medication
Bid	6.73	1.29	0.59	0.93
Expected Position	1	3	3	2

Measuring Conversions

Conversion > Phone calls, emails or conversion pages (forms)

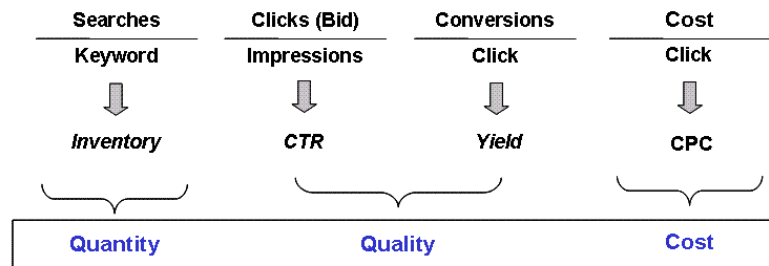


Key Metrics

- ◆ Key Levers
- ◆ Organic traffic:
 - Metrics: monitor number of pages indexed (benchmark against competitors and look at trends), pagerank (changes slowly), weekly/monthly rank on search engines for your keyword universe (avg, weighted avg pos)
- ◆ Paid traffic?

Key Metrics

Metrics: Impressions Clicks Conversions Cost



Key Metrics

- ◆ Cost Per Click (CPC) or Per Conversion by Search Engine
 - These should be similar
- ◆ Conversion Quality by Search Engines
 - Determines bid intensity
- ◆ Click/conversion Volume Generated
 - Break down by source
- ◆ Average Position (Rank) by Category
 - Align with marketing goals
 - » Branding (firm level)
 - » Promoting specific prod lines
 - » LTV of customer
 - » Revenues
 - » Margins

Sponsored Search with Contexts

Sponsored Search with Contexts

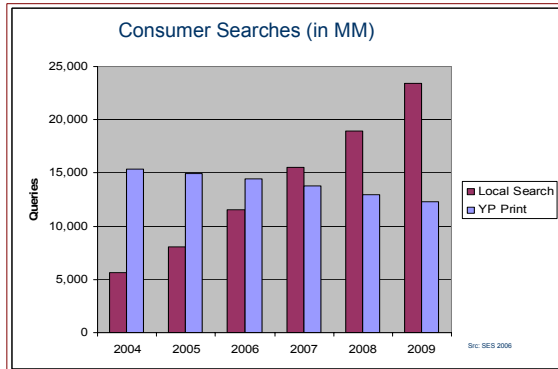
- ◆ Context: “any auxillary information that might accompany a search, and might include information that is factual, estimated or inferred. Natural examples of contexts include the zip code, gender, or abstract “intentions”...” (Even-dar et al. 2007)
- ◆ Real-world examples
 - Bid on “dentist” but have different bids for different regions (identified by IP addresses)
 - Retailer bids on clothing but places higher premium on women under 30

Context-based Auctions: Market Opportunity

Local Advertising Market

\$ 100 B

220 M



Online searches ↑ 50% pa

Examples

Welcome liza, to your Natpal Account
Not Liza? [Click Here](#)

1. Search 2. Site 3. Stats Overview **Get Found.**

Natpal LIVE Login >> Password >> **Go!** [Forgot Password?](#)

Pending Account Tab

Welcome Liza,
Thanks for setting up your Natpal Live Account. If you need any assistance please email support@natpal.com or call 800.4Natpal.

Segment:
Dentist

Target Area:
Philadelphia

Budget:
\$530.00

Advert Type:
dentistdc.com

Content:
Entered

Features:
Entered

Billing Info:
Not Entered

Select Segment >>

Electrician

Auto
Bakery
Banquet Room
Bridal Shop
Caterer
Chiropractic
Computer Repair
Dentist
Dermatologist
Ecommerce Website
Electrician

Select Categories >>

Residential
 Commercial
 100 & 200 AMP Services
 Air Conditioning Lines
 Equipment Installation
 New Construction Pre-Wiring
 Remodeling Re-Wiring
 Indoor & Outdoor Lighting
 Computer & Telephone Cabling
 Computer & UPG Power Wiring
 Security and Fire Alarms

Account Info [Go!](#) **Save**

[Home](#) | [Case Studies](#) | [Corporate](#) | [Contact Us](#) | [Partners](#) | [Support](#)

Select Context

Natpal Get Found.™

Welcome Liza, to your Natpal Account
Not Liza? [Click Here](#)

1. Search 2. Site 3. Stats Overview **Get Found.**

Natpal LIVE Login >> Password >> **Go!** Forgot Password?

Pending Account Tab

Welcome Liza,
Thanks for setting up your Natpal Live Account. If you need any assistance please email support@natpal.com, or call 800.4NATpal.

Segment:
Dentist

Target Area:
Philadelphia

Budget:
\$530.00

Advert Type:
dentistdc.com

Content:
Entered

Features:
Entered

Billing Info:
Not Entered

Select Your Area >>



Account Info: [Edit](#)

Save

[Home](#) | [Case Studies](#) | [Corporate](#) | [Contact Us](#) | [Partners](#) | [Support](#)

Select Budget

Natpal Get Found.™

Welcome Liza, to your Natpal Account
Not Liza? [Click Here](#)

1. Search 2. Site 3. Stats Overview **Get Found.**

Natpal LIVE Login >> Password >> **Go!** Forgot Password?

Pending Account Tab

Welcome Liza,
Thanks for setting up your Natpal Live Account. If you need any assistance please email support@natpal.com, or call 800.4NATpal.

Segment:
Dentist

Target Area:
Philadelphia

Budget:
\$530.00

Advert Type:
dentistdc.com

Content:
Entered

Features:
Entered

Billing Info:
Not Entered

Enter Your Budget >>

Natpal Budget & Traffic Estimator



Slide Cursor To Change Budget

Estimate= 20 Calls | \$530

Save

[Home](#) | [Case Studies](#) | [Corporate](#) | [Contact Us](#) | [Partners](#) | [Support](#)

Keyword Creation

**Automatically
Configures»**

Keywords Created

construction electrician philadelphia pa
electrician in philadelphia
construction electricians philadelphia
electrical contractor website in philadelphia
commercial electricians in philadelphia
construction electrician in philadelphia
certified electricians philadelphia pa
commercial electricians in philadelphia
electrical utility contractor philadelphia
commercial electricians philadelphia
electric help for systems philadelphia pa
construction electricians philadelphia pa
certified electricians philadelphia pennsylvania
certified electrician philadelphia
electrical contractor website philadelphia
electric help in philadelphia

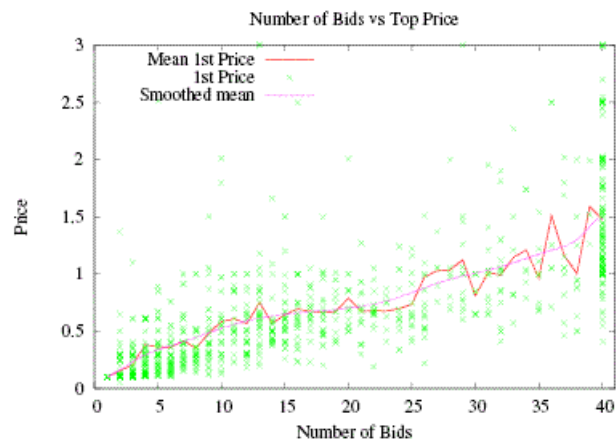
Internet and Local

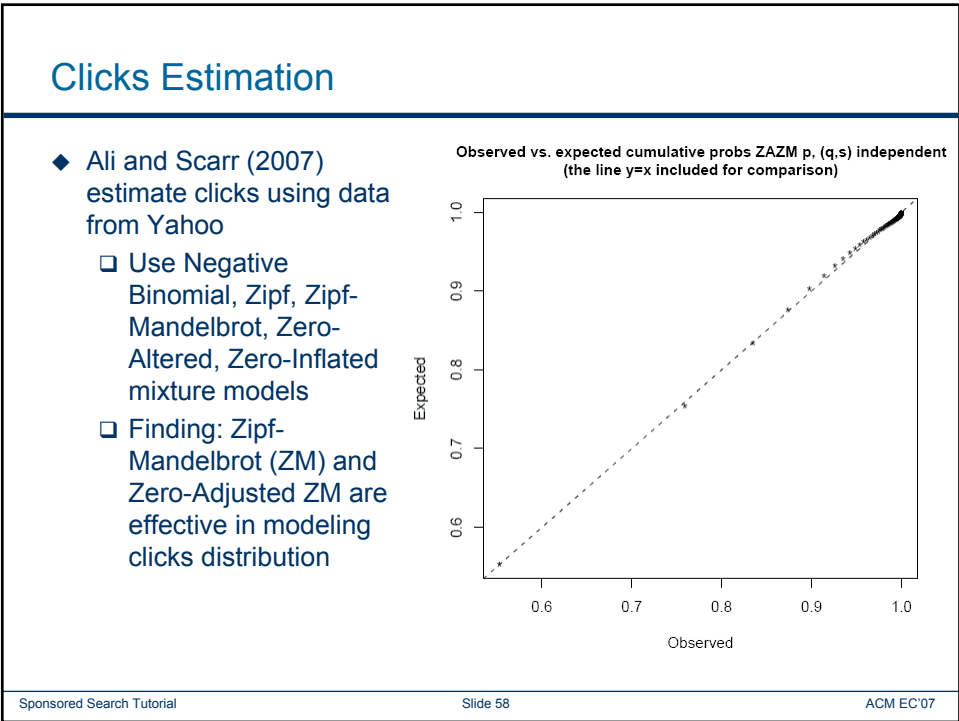
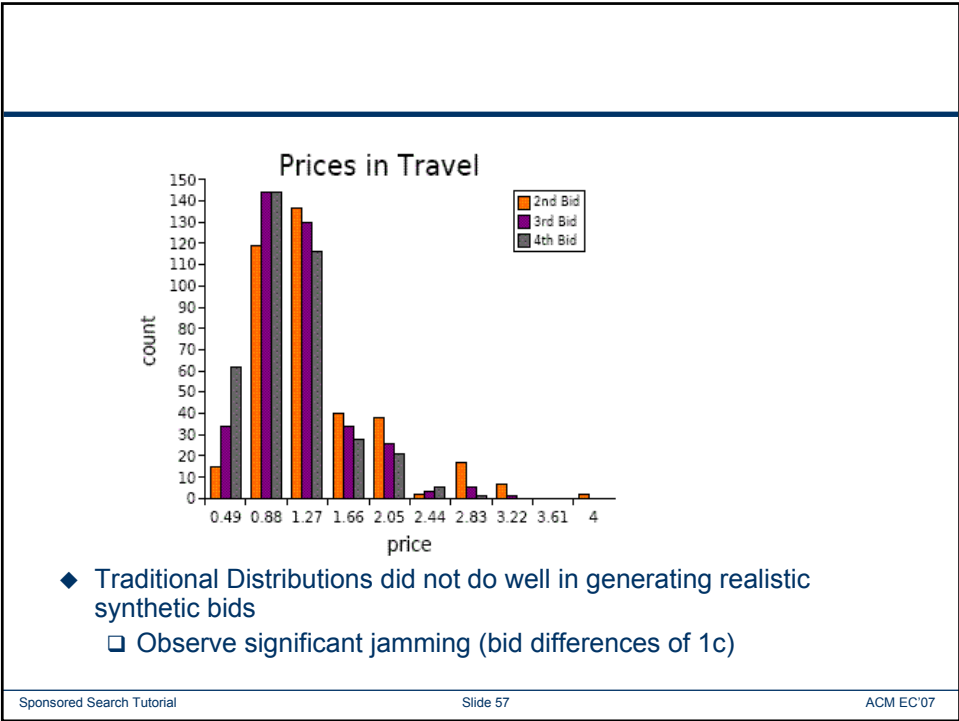
- ◆ Results from Even-Dar et al. (2007): Overall social welfare (advertiser and auctioneer surplus) increases when moving from standard to context-based auctions
 - However, one party can benefit at the expense of the other
- ◆ Broader Implications of Context-based Auctions
 - YP market is fundamentally changing
 - Many YPs have launched online YPs (superpages.com)
 - » But search driving most of the traffic

Empirical Issues in Sponsored Search

Synthetic Bid Generation

- ◆ Ganchev et al. (2007) study bids placed in Yahoo's (old) open auction





Estimating CTR for new Ads

- ◆ Search engine needs to estimate CTR of new ad in order to place it
- ◆ Hard to estimate in the absence of historical data
- ◆ Richardson et al. (2007) run a logistic regression to estimate probability of click
 - Factors
 - » CTR for other ads (for same keyword)
 - » CTR for related terms (subset of superset of terms)
 - » Ad Quality (appropriate capitalization, use of punctuation, etc)
 - Each factor helped improve fit by 20% or more

Conclusions

- ◆ Several Challenges and Open Questions
 - Mechanism Design
 - » What are the best rules for ranking, payment
 - » Open versus opaque auction
 - Bid Optimization
 - » Combinatorial Problem
 - » Stochastic
 - Keyphrase Generation
 - » Machine Learning
 - Empirical Studies
 - » Behavior Modeling: Estimating clicks, conversions
 - » Estimating clicks and other functional relationships

Conclusions

- ◆ Search Engine Marketing is an interdisciplinary field
 - Economics (Mechanism Design)
 - Computer Science (Algorithms, Machine Learning)
 - Statistics (Resolving uncertainty)
 - Marketing (Modeling Consumer Behavior)

References

- ◆ Ali, K. and Scarr, M. 2007. Robust methodologies for modeling web click distributions. In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007).
- ◆ Deeparnab Chakrabarty, Yunhong Zhou and Rajan Lukose "Budget constrained bidding in Keyword Auctions and Online Knapsack problems", Third workshop on Sponsored Search, Banff 2007.
- ◆ E. Even-Dar, J. Wortman and M. Kearns, "Sponsored Search with Contexts", Third Workshop on Sponsored Search Auctions, WWW 2007.
- ◆ K. Ganchev, A. Kulesza, J. Tan, R. Gabbard, Q. Liu. Empirical Price Modeling for Sponsored Search. Third Workshop on Sponsored Search Auctions, WWW 2007.
- ◆ Holthausen, D and G. Assmus, "Advertising Budget Allocation Under Uncertainty," Management Science, 28 (May 1982), 487-499.
- ◆ K. Hosanagar and P. Stavrinides, "Optimal Bidding in Online Slot Auctions", First International Symposium of Information Systems, Indian School of Business, Hyderabad, India.

References

- ◆ A. Joshi and R. Motwani, "Keyword Generation for Search Engine Advertising", Workshop Proceedings of IEEE ICDM 2006, December 2006
- ◆ Muthukrishnan, S.; Pal, Martin; Svitkina, Zoya, "Stochastic Models for Budget Optimization in Search-Based Advertising ", Third workshop on Sponsored Search, Banff 2007.
- ◆ Richardson, M., Dominowska, E., and Rago, R. 2007. Predicting clicks: estimating the click-through rate for new ads. In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007).
- ◆ P. Rusmevichientong and D. P. Williamson, "An Adaptive Algorithm for Selecting Profitable Keywords for Search-Based Advertising Services," Proceedings of the Seventh ACM Conference on Electronic Commerce, pp. 260-269, 2006.