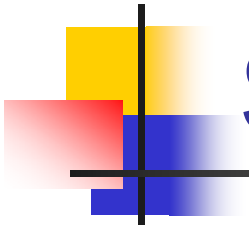
The Economics of Internet Search

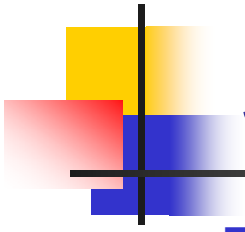
Hal R. Varian

Sept 31, 2007



Search engine use

- Search engines are very popular
  - 84% of Internet users have used a search engine
  - 56% of Internet users use search engines on a given day
- They are also highly profitable
  - Revenue comes from selling ads related to queries



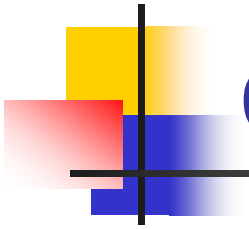
Search engine ads

- Ads are highly effective due to high relevance
  - But even so, advertising still requires scale
    - 2% of ads might get clicks
    - 2% of clicks might convert
    - So only .4 out a thousand who see an ad actually buy
    - Price per impression or click will not be large
    - But this performance is good compared to conventional advertising!
- Search technology exhibits increasing returns to scale
  - High fixed costs for infrastructure, low marginal costs for serving

Summary of industry economies

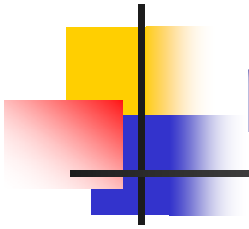
- Entry costs (at a profitable scale) are large due to fixed costs
- User switching costs are low
  - 56% of search engine users use more than one
- Advertisers follow the eyeballs
  - Place ads wherever there are sufficient users, no exclusivity
- Hence market is structure is likely to be
  - A few large search engines in each language/country group
  - Highly contestable market for users
  - No demand-side network effects that drive towards a single supplier so multiple players can co-exist

What services do search engines provide?

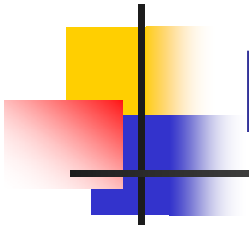


- Google as yenta (matchmaker)
  - Matches up those seeking info to those having info
  - Matches up buyers with sellers
- Relevant literature
  - Information science: information retrieval
  - Economics: assignment problem

Brief history of information retrieval

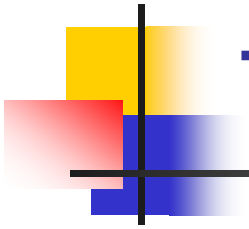


- Started in 1970s, basically matching terms in query to those in document
- Was pretty mature by 1990s
- DARPA started Text Retrieval Conference
  - Offered training set of query-relevant document pairs
  - Offered challenge set of queries and documents
  - Roughly 30 research teams participated



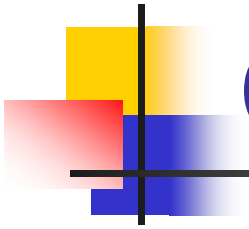
Example of IR algorithm

- Prob(document relevant) = some function of characteristics of document and query
  - E.g., logistic regression $p_i = X_i \beta$
- Explanatory variables
  - Terms in common
  - Query length
  - Collection size
  - Frequency of occurrence of term in document
  - Frequency of occurrence of term in collection
  - Rarity of term in collection



The advent of the web

- By mid-1990s algorithms were very mature
- Then the Web came along
  - IR researchers were slow to react
  - CS researchers were quick to react
- Link structure of Web became new explanatory variable
  - PageRank = measure of how many important sites link to a given site
  - Improved relevance of search results dramatically

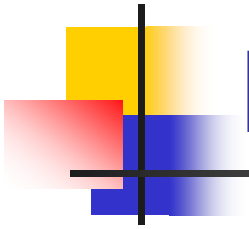


Google

- Brin and Page tried to sell algorithm to Yahoo for $1 million (they wouldn’t buy)
- Formed Google with no real idea of how they would make money
- Put a lot of effort into improving algorithm

Why online business are different

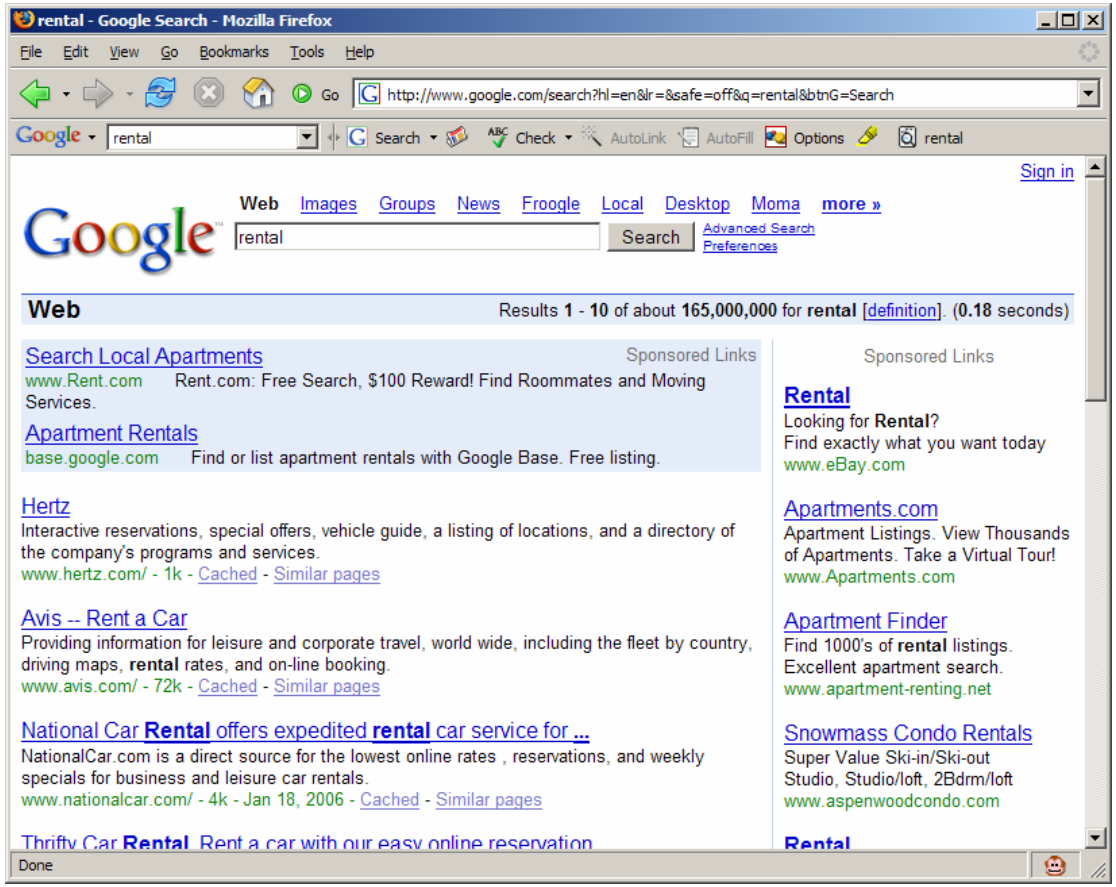
- Online businesses (Amazon, eBay, Google...) can continually experiment
  - Japanese term: *kaizen* = “continuous improvement”
  - Hard to really do continuously for offline companies
    - Manufacturing
    - Services
  - Very easy to do online
    - Leads to very rapid (and subtle) improvement
    - Learning-by-doing leads to significant competitive advantage



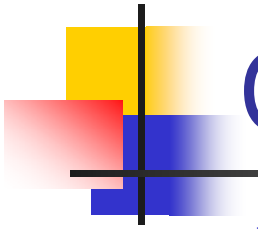
Business model

- Ad Auction
  - GoTo's model was to auction search results
  - Changed name to Overture, auctioned ads
  - Google liked the idea of an ad auction and set out to improve on Overture's model
- Original Overture model
  - Rank ads by bids
  - Ads assigned to slots depending on bids
    - Highest bidders get better (higher up) slots
  - High bidder pays what he bid (1st price auction)

Search engine ads

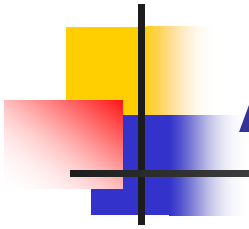


- Ads are shown based on query+keywords
- Ranking of ads based on expected revenue



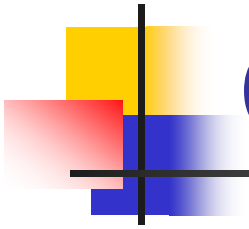
Google auction

- Rank ads by bid x expected clicks
  - Price per click x clicks per impr = price per impression
  - Why this makes sense: revenue = price x quantity
- Each bidder pays price determined by bidder below him
  - Price = minimum price necessary to retain position
  - Motivated by engineering, not economics
- Overture (now owned by Yahoo)
  - Adopted 2nd price model
  - Currently moving to improved ranking method



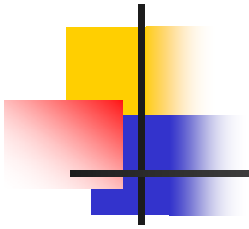
Alternative ad auction

- In current model, optimal bid depends on what others are bidding
- Vickrey-Clarke-Groves (VCG) pricing
  - Rank ads in same way
  - Charge each advertiser cost that he imposes on other advertisers
  - Turns out that optimal bid is true value, no matter what others are bidding



Google and game theory

- It is fairly straightforward to calculate Nash equilibrium of Google auction
  - Basic principle: in equilibrium each bidder prefers the position he is in to any other position
  - Gives set of inequalities that can be analyzed to describe equilibrium
  - Inequalities can also be inverted to give values as a function of bids



Implications of analysis

- Basic result: *incremental cost per click has to be increasing in the click through rate.*
- Why? If incremental cost per click ever decreased, then someone bought expensive clicks and passed up cheap ones.
- Similar to classic competitive pricing
  - Price = marginal cost
  - Marginal cost has to be increasing

Simple example

- Suppose all advertisers have same value for click v
  - Case 1: Undersold auctions. There are more slots on page than bidders.
  - Case 2: Oversold auctions. There are more bidders than slots on page.
- Reserve price
  - Case 1: The minimum price per click is (say) p_m (~ 5 cents).
  - Case 2: Last bidder pays price determined by 1st excluded bidder.

Undersold pages

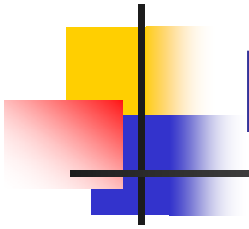
- Bidder in each slot must be indifferent to being in last slot

$$(v - p_s)x_s = (v - r)x_m$$

- Or

$$p_s x_s = v(x_s - x_m) + rx_m$$

- Payment for slot s = payment for last position + value of incremental clicks

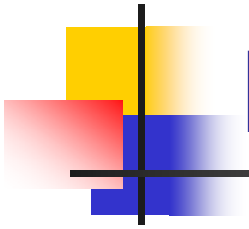


Example of undersold case

- Two slots
  - x_1 = 100 clicks
  - x_2 = 80 clicks
  - v=50
  - r=.05
- Solve equation
  - p_1 100 = .50 x 20 + .05 x 80
  - p_1 = 14 cents, p_2=5 cents
  - Revenue = .14 x 100 + .05 x 80 = $18

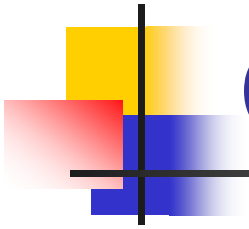
Oversold pages

- Each bidder has to be indifferent between having his slot and not being shown:
- So $$(v - p_s)x_s = 0$$
$$p_s = v$$
- For previous 2-slot example, with 3 bidders, p_s=50 cents and revenue = .50 x 180 = $90
- Revenue takes big jump when advertisers have to compete for slots!



Number of ads shown

- Show more ads
  - Pushes revenue up, particularly moving from underold to oversold
- Show more ads
  - Relevancy goes down
  - Users click less in future
- Optimal choice
  - Depends on balancing short run profit against long run goals



Other form of online ads

- Contextual ads
  - AdSense puts relevant text ads next to content
  - Advertiser puts some Javascript on page and shares in revenue from ad clicks
- Display ads
  - Advertiser negotiates with publisher for CPM (price) and impressions
  - Ad server (e.g. Doubleclick) serves up ads to pub server
- Ad effectiveness
  - Increase reach
  - Target frequency
  - Privacy issues

Conclusion

- Marketing as the new finance
- Availability of real time data allows for fine tuning, constant improvement
- Market prices reflect value
- Quantitative methods are very valuable
- We are just at the beginning...