


**NBA 600**  
**Searching the Web and Beyond**  
 Class 9, Mon 11/ 12

Prof. Dan Huttenlocher

### Administrative


- Reminder: short papers due on Wednesday
- Second quiz today



2

### Google Business Overview


- Main business is advertising supported Web search – AdWords
  - Customers seeking information enter queries
  - Advertisers bid on terms and pay per click
  - Relevant ads shown along side search results
- Large and increasing business of ads on third-party sites – AdSense
  - Relevant ads displayed on Web pages, also Google search box on Web pages
  - Advertisers pay per click, site publisher gets cut



3

### Google Financials


- Revenue of 3Q07 \$4.23B
  - 59% y-o-y growth, down from 73% prev. yr.
  - Net income \$1.07B
- Revenue breakdown
  - \$2.73B from Google sites
    - Largely AdWords search keywords
    - 68% y-o-y growth
  - \$1.45B from third party sites
    - AdSense site placement ads
    - 40% y-o-y growth



4

### What a Search Engine Does


- Crawler or spider finds pages on the Web
  - By following hyperlinks from pages already found
    - Like exploring from a starting town, keeping track of each unexplored neighboring town and eventually returning there
- Indexer stores which terms are relevant to which pages
  - Like the index in the back of a book
- Ranking algorithm determines which pages most relevant to given query



5

### Big Advance of Google

- Google quickly displaced AltaVista between 2000 and 2003 because of better search results
- Largely due to PageRank which takes link structure into account as well as textual content
  - Links are a form of citation of a page
    - More citations means more important
  - Pages that are themselves important convey even more importance
    - Self-reinforcing or “rich get richer”



6

## Constant Tuning

- Search engine optimization companies work to raise rankings of sites in listings
  - Sometimes “tricking” by adding many dummy pages with links
- Search engines constantly tune their ranking methods
  - Balancing link structure, text content, rate of change, degree of linking
  - Often has substantial effect on rankings
    - E.g., rise of blogs, which have high degree of linking

## Search in the Value Chain

- Search based on keywords or descriptions
  - Part of many activities
    - Can view many transactions as composed of: search, payment, fulfillment
    - News, sports and entertainment often involve substantial search component
- People generally don't view search as goal
  - In service of something else
- Potential of search to influence actions
  - But perceived bias in search results can have negative impact on value to user

## Variants on Search

- About and other sites that provide more than automated keyword searches
- New entrants that try to combine keyword search and user annotation (tagging)
- User annotation at specific sites such as Flickr, Youtube
- Image search
- Various niche search facilities, including airfare searching

## Competitive Landscape in Search

- Barriers to entry
  - Getting higher due to scale of data and servers needed (hundreds of thousands at Google)
- Substitutes
  - Potential for smarter tools that better support end goal, but not clear will be as useful
- Customer power
  - Low switching costs, little lock-in or loyalty
- Supplier power
  - Content providers? Can they effectively opt out?

## Revenue Models for Search

- For many years was a challenge
  - Paid inclusion allows sponsors to have their pages listed in response to certain queries
    - Can lead to lowering of search quality, customer dissatisfaction
    - FTC investigations about un-credited paid inclusion being misleading
- Google always argued against paid inclusion advertising – diminishes value
  - Sought ways of having “relevant ads” appear adjacent to search results

## Bidding for Terms

- Overture launched a market for search terms
  - Advertisers bid a per-click fee for being listed for certain words or phrases
    - Placement ordered by bid
    - Bidding gave advertisers “pricing context”
- Google still sought mechanism for assessing relevance of ads, not just bids
  - Focused on useful information even in ads
  - Highest bidder often not relevant content

## AdWords

- Google's mechanism of bidding for listings
  - Placement based on product of click through (CTR) and cost-per-click (CPC)
    - More relevant items tend to be placed higher due to more click-throughs
  - Advertisers bid a maximum CPC they are willing to pay
    - Actual fee is lowest amount that is \$.01 higher than next lowest ranked bidder
      - Auction theory, second price auctions
    - Can set daily expenditure budget, as hard to predict precise number of clicks and cost

## Key Benefits of AdWords Model

- Ads that users find relevant
  - Benefits end-user by supplementing search results with valuable sponsored links
  - Benefits advertisers by reaching likely buyers
- Costs reflect value both to end-users and advertisers
  - Bidding for terms
- Costs bounded and predictable
  - Not only bid for terms but also set explicit budgets

## Changing Views of Advertising

- Price only one component of when ads displayed
  - Has to be relevant as measured by user actions (clicks)
- Initially resulted from focus on user value
  - But also valuable to advertisers in targeting their campaigns
    - Quickly learn which ad copy and which keywords are effective
    - Start with large number of terms, automatically shown CTR and can focus appropriately

## AdSense

- Google's technological advantage is in determining what's what on the Web
  - Combination of content and link analysis
- Allows Google to do good job of selecting ads for Web content not just search
  - E.g., distinguish different meanings of terms, potentially different viewpoints
- AdSense created in early 2003 to do this
  - Web page publishers can place ads as well as Google search box on their pages
    - Get paid by Google for clicks

## Localized Advertising

- Online ads can be restricted to run only in certain geographies
  - Google now offers about 200 regions in US corresponding to traditional TV markets
  - Improving resolution globally, as non-US revenue 40% in 3Q06
- Based on services that attempt to determine geography from IP address
  - Not completely accurate
    - Some large service providers such as AOL/Time Warner particularly problematic

## Localized Search

- Based on combination of search terms and a user-specified location
  - Rather than IP address as for ad targeting
- Google experimented with searching only Web pages but results disappointing
  - Now combine with directory search and maps
- Industry estimates are 20-30% of current searches local in nature
  - Not previously well served by searches of entire Web

### **Looming Changes in Competition**

- Microsoft moving into online services and Google moving into desktop applications
  - Low switching costs for search applications
- Search term ads as substitute for Ebay listings
  - Both can be form of “classified ad”
- Search as intermediary in online retail
  - Possibly directing users to sites other than dominant brands such as Amazon

### **Next Time**

- Online communities and social networking