Introduction to computational advertising

Bogdan Cautis

Outline

- From IR to IS
- Advertising on the Web
  - The evolution of Web advertising
  - Terminology
  - History
- Advertising setting and problems
  - Display ads
  - Sponsored search
  - Content match
- The economics of Web advertising
- Conclusion

From information retrieval to information supply

A brief history of Web search

- Based mainly on Information Retrieval (IR)
  - IR as a discipline has been around for 50-60 years
- June 11, 1994 – Brian Pinkerton: WebCrawler
- Dec 15, 1995 – AltaVista (crawled at 2.5 M pages per day, had 30 M pages)
- 1995, Yahoo (Yet Another Hierarchical Officious Oracle)
- 1998 – Google (googol 10^{100}), 2004 – Google IPO

Today

- «Without search engines the Web would not exist »
  (Andrei Broder, VP Emerging Search Technologies at Yahoo)
- Both a technology artifact and a social environment
- The normal way of life, search interaction is accepted paradigm
- Useful? For instance, it makes aggregation possible
  - Unlimited “selection stores” possible: Amazon

IR basics

- Corpus: Fixed document collection
- Goal: Retrieve documents with information content that is relevant to user’s information need
- Relevance
  - For each query Q and stored document D in a given corpus assume there exists relevance Score(Q, D)
- **The context is ignored**
- **The users are ignored**

User needs

- Informational
  - want to learn about something (~40% / 65%)
- Navigational
  - want to go to that page (~25% / 15%)
- Transactional
  - want to do something (web-mediated) (~35% / 20%)

    - access a service

    - download

    - shop

Search engines generations

- First generation -- use only “on page”, text data
  - word frequency, language
- Second generation -- use off-page, web-specific data
  - link (or connectivity) analysis
  - sophisticated mathematical methods
  - click-through data (What results people click on)
  - anchor-text (How people refer to this page)
- Third generation -- answer “the need behind the query”
  - focus on user need, rather than on query
  - semantic analysis -- what is this about?
  - integrates multiple sources of data, context
  - **help the user!**

Third generation search engines

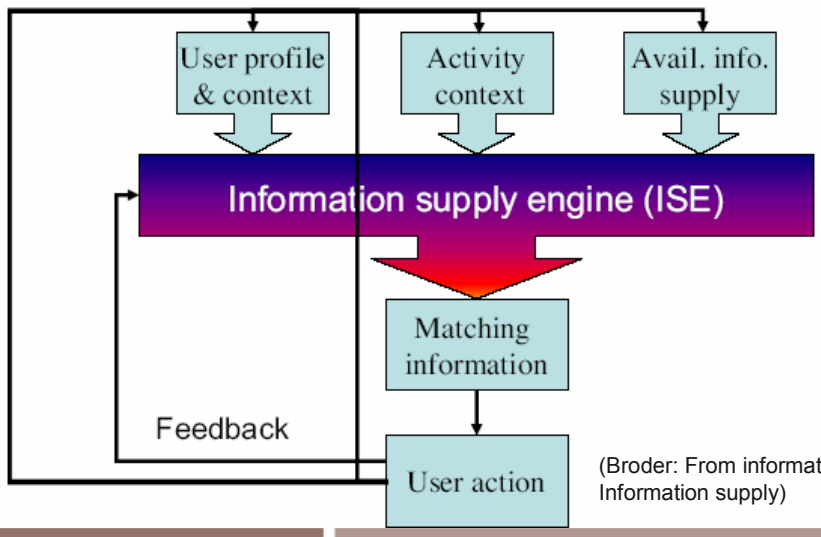
Understanding "the need behind the query" rather than simply returning query matches is

Triple win

- A win for users (better results)
- A win for content providers (focus)
- A win for search engines (monetization)

Information Supply

- **From information retrieval to information supply: Move from syntactic matching to semantic matching**



(Broder: From information retrieval to Information supply)

Examples

- Subscriptions (e-mail, RSS, etc)
- Alerts – News
- E-commerce sites: accessories, commentaries, related purchases, etc
- Automatic annotations
- Recommendations
- Contextual ads & search driven ads
- – …

Web advertising

Introduction

- **The Internet advertising spending is estimated over 17 billion dollars in 2006 ($150B total).**
- **97% of search revenue**
  - broadband is cheap, ubiquitous
  - “getting things done” easier on the Internet
- **Why does it work**
  - massive scale, automated
  - key: monetize more and better, “learn from the data”
  - new discipline: “Computational advertising”

Computational advertising

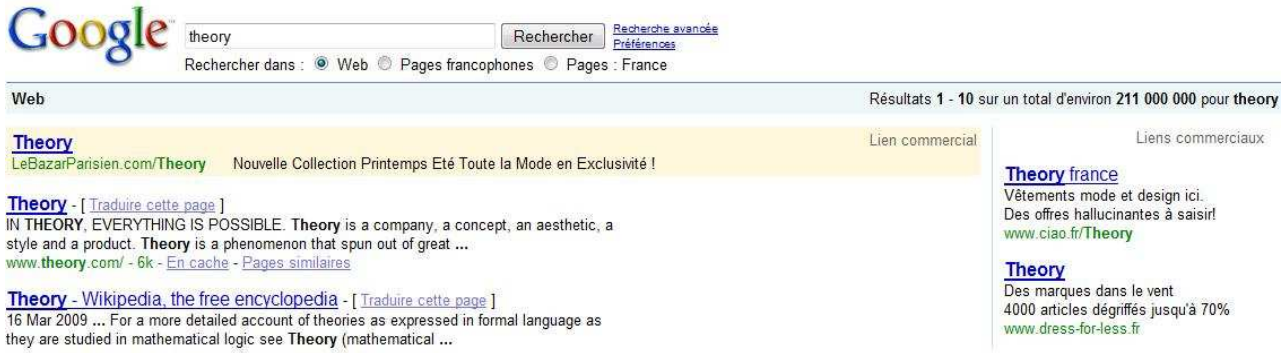
- **New scientific sub-discipline, at the intersection of**
  - large scale search and text analysis
  - information retrieval
  - statistical modeling
  - machine learning
  - optimization
  - microeconomics

Ad types

- **Three main types of textual Web advertising:**
  - ***Sponsored search*** which serves ads in response to search queries
  - ***Content match*** which places ads on third-party pages
  - ***Display advertising*** (banner ads)
- **Ads are information!**

A sponsored search ad

Another example



A content match ad (1)

(Broder: From information retrieval to Information supply)

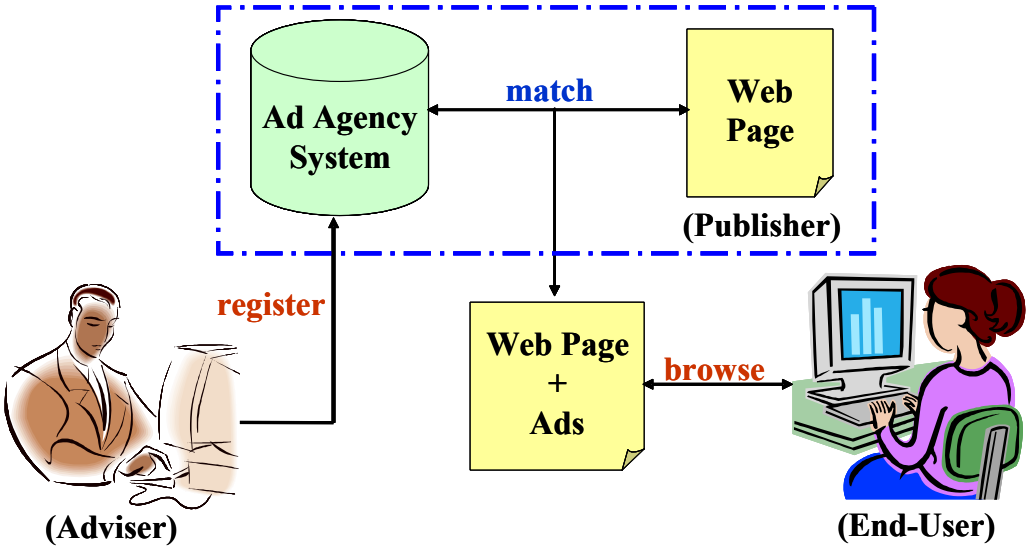
A content match ad (2)

(Broder: From information retrieval to Information supply)

Contextual Advertising Basic

- **Four interactive entities:**
  - The **publisher** is the owner of Web pages on which advertising is displayed.
  - The **advertiser** provides the supply of ads.
  - The **ad network** is a mediator between the advertiser and the publisher, who selects the ads that are put on the pages.
  - **End-users** visit the Web pages of the publisher and interact with the ads.

Overview of Ad display



Behind the curtains (sponsored search)

- Manual or automated review process to ensure that advertiser content is in fact relevant to the target keyword
- Matching advertiser content to user queries as they are received
- Displaying advertiser content in some rank order
- Gathering data, mesuring clicks, charging advertisers based on consumer clicks, etc.

Different revenue flavors

Revenue models

- CPM: Cost Per iMpression
- CPC: Cost Per Click
- CPV: Cost Per Visitor
- CPA: Cost Per Activity

Revenue models

- Example: suppose we show an ad N times on the same spot
- Under CPM: Revenue = N * CPM
- Under CPC: Revenue = N * CTR * CPC

Depends on the auction mechanism

Click-through rate

(probability of a click given an impression)

Revenue models

- Example: suppose we show an ad N times on the same spot
- Under CPM: Revenue = N * CPM
- Under CPC: Revenue = N * CTR * CPC
- Under CPA: Revenue = N * CTR * Conv. Rate * CPA

Conversion rate:

(probability of a user conversion on the advertiser's landing page given a a click)

Revenue models

Revenue dependence:

- CPM: website traffic
- CPC: + ad relevance
- CPA: + landing page quality

- From 1st to 3rd: more relevant for advertisers, bigger prices and bids!

Web advertising history

INSTITUT TELECOM

How it all started

- **In 1998, GoTo, later Overture Systems**
  - sponsored search
  - Yahoo acquired Overture in 2003
- **BeFirst followed in 1999**
- **Google adopted the model**
  - added click feedback (2002)
- **MSN extended the model to behavioral targeting**
- **HotWired introduced banner advertising in 1994**
- **The search engine OpenText offered preferred listings, in 1996.**

INSTITUT TELECOM

How it all started (2)

- First, CPM based (a.k.a. cost per mille)
- In 1996, Yahoo agreeed to charge Procter&Gamble by CPC
- By 1997, DoubleClick was offering CPA pricing
- In 2003, Overture, Google and FindWhat introduced automated tools for measuring CPA
- Auction models: in 1997, FlyCast and Narrowline
- Google's (2002) generalized second-price auction (GSP) in AdWords
- Pay the bid of the next highest bidder
- Later adopted by Yahoo!/Overture and others

Advertising settings and problems

Athens week, March 2009

INSTITUT TELECOM

Core issues

- What do you show to a user?
- How does the user interact with the ad system?

Display ads

- Just pick ads
  - graphically displayed
  - mostly for brandness awareness
  - revenue model is typically CPM

Display ads

- **Traditional advertising model:**
  - Ads are targeted at particular demographics
    - GM adds on Yahoo autos shown to « males above 55 »
    - Mortgage ad shown to « everybody Yahoo Front page »
  - Book a slot well in advance
    - « 2M impressions in Jan next year »
    - impressions guaranteed by the ad network!

Display ads

- **Fundamental problem: guarantee impressions to advertisers**
  - predict supply
    - how many impressions available
    - demographics overlap
  - predict demand
    - how much will advertisers want each demographic
  - find the « optimal » allocation
- **Forecast accuracy is critical!**

Content match and sponsored search ads

Content match

- **Pick adds by matching them to content**
- **The user intent is unclear**
- **Webpage can be big and noisy**

Sponsored search

- **Given a search query**
- **Pick adds by matching them to the query**
- **User declares her intention**
- **Query is short and less noisy than Content Match**

The main issues

- **Given a « query »**
  - Select the top-k ads to be shown on the k slots in order to maximize total expected revenue
- **What affects the total revenue**
  - Relevance of the ad to the query
  - Bids on the ads
  - User experience on the landing page (ad « quality »)

Ad relevance computation

- **IR based**
  - use a search engine to match ads to context
    - ads are the « documents »
    - Context (user query or webpage content) are the query
  - Problem: word matches might not always work
  - Need to extract topical information
- **Machine learning from clicks**
  - Estimate CTR=Pr(click | ad, query, user)
  - Ad-ad similarity & collaborative filtering

Collaborative filtering connection

- **CTR feedback for each (query, ad) pair**
  - learn the « best match between a user in a given context and a suitable advertisement »
- **data is sparse, in order to get the best match, we need to find similar ads, pages, and users.**

Ad relavance by online learning

- **Online learning**
  - previous approaches learned from historical data
  - slow response to emerging patterns,
  - initial biases never corrected

  if the system never showed “golf classes” for “iPod” it can never learn if this matching is good.

- **Solution: combine exploitation with exploration**
  - pick ads that are good according to current model
  - pick ads that increase your knowledge about the entire space of ads.

Online content matching

- **Web advertising for two types of Web page:**
  - ***Static page* (Offline)**: the matching of ads can be based on prior analysis of their entire content.
  - ***Dynamic page* (Online)**: ads need to be matched to the page *while it is being served to the end-user*. Thus, limiting the amount of time allotted for its content analysis.

Online content matching

- When a user views a page, the ad selection engine has only a couple hundred milliseconds to provide the ads.
- Offline approach: works well for static content pages that are displayed repeatedly.
- Ads need to be matched to the page while it is being served to the end-user, with extremely limited time for content analysis.

The economics of Web advertising

Auctions

- **Generalized First-Price Auctions 1997 auction revolution by Overture (then GoTo)**
- **Pay per-click for a particular keyword**
- **Links arranged in descending order of bids**
- **Pay your bid**

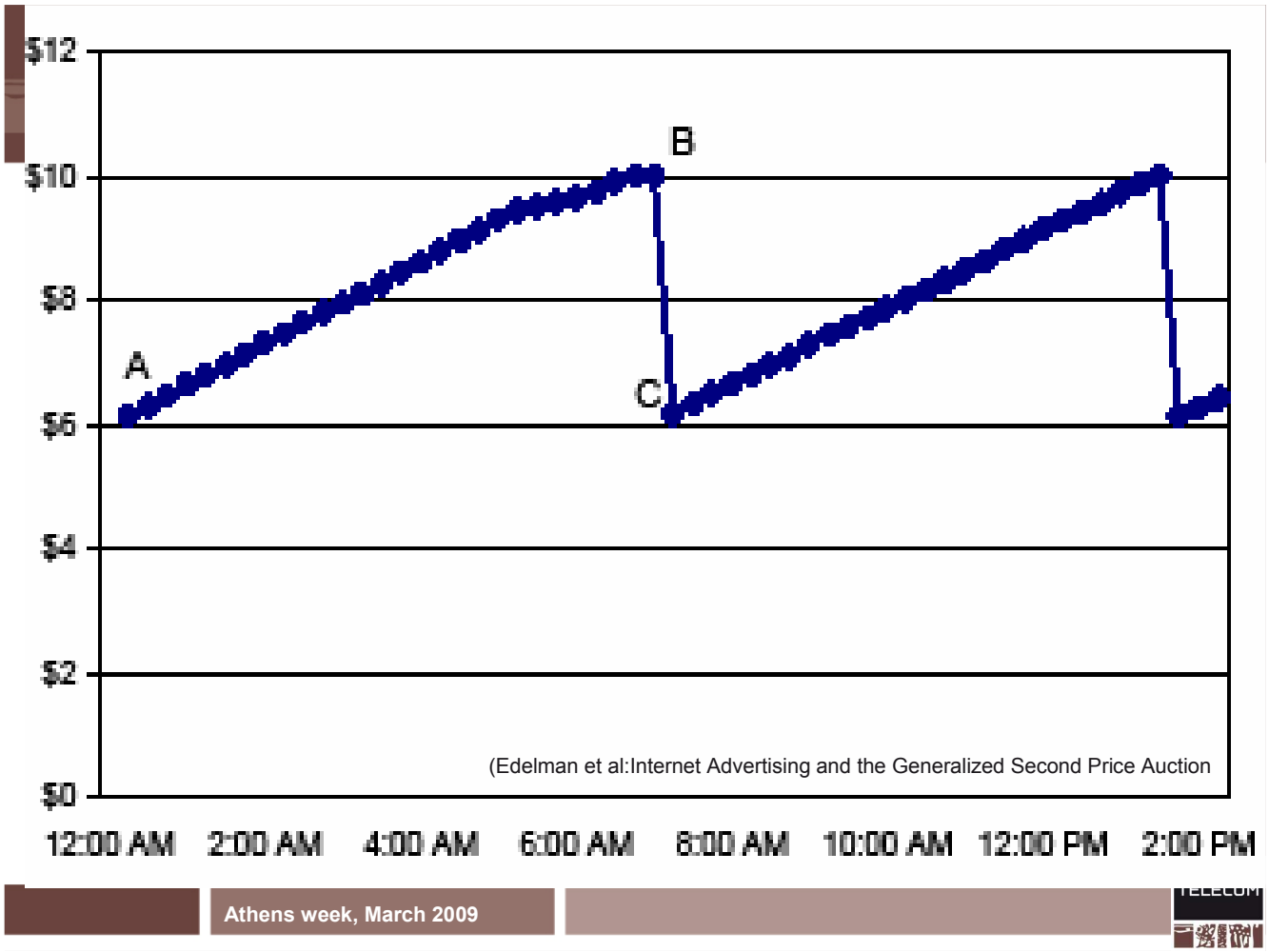
Problem: GFP is unstable because bids can be adjusted dynamically

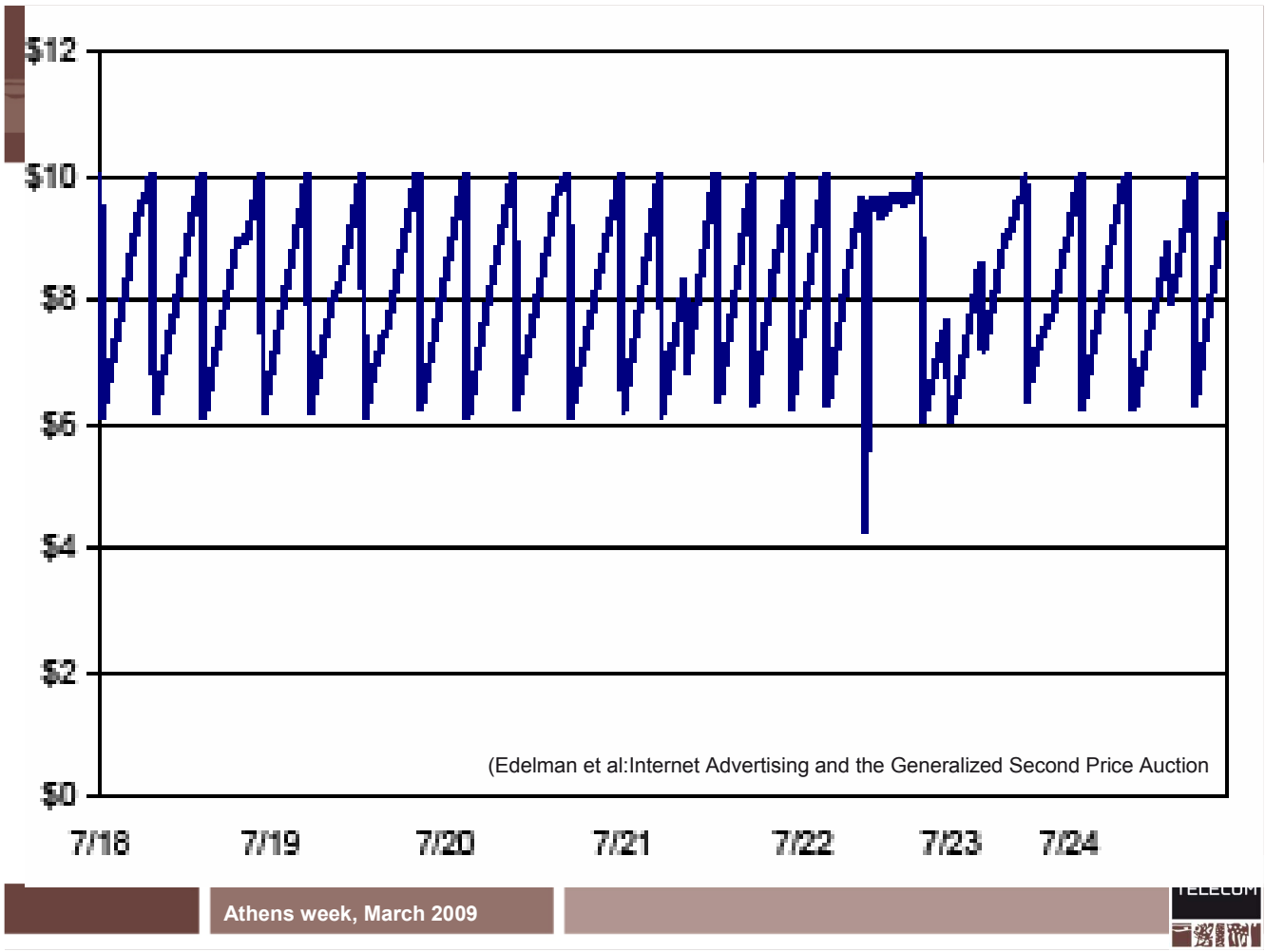
Example on GFP

Two slots and three bidders.

- ad in first slot: 200 clicks per hour,
- ad in second slot: 100.

- bidders 1, 2, and 3 have values per: $10, $4, and $2
- if bidder 2 bids $2.01, to make sure he gets a slot.
- bidder 1 will not want to bid more than $2.02
- bidder 1 gets the top spot, but then bidder 2 will want to revise his bid to $2.03 to get the top spot,
- bidder 1 will in turn raise his bid to $2.04, and so on.


$12
$10
$8
$6
$4
$2
$0
A
B
C
(Edelman et al:Internet Advertising and the Generalized Second Price Auction
12:00 AM
2:00 AM
4:00 AM
6:00 AM
8:00 AM
10:00 AM
12:00 PM
2:00 PM
Athens week, March 2009


$12
$10
$8
$6
$4
$2
$0
(Edelman et al:Internet Advertising and the Generalized Second Price Auction
7/18
7/19
7/20
7/21
7/22
7/23
7/24
Athens week, March 2009

Generalized second-price auctions

- **A bidder in position i will never want to pay more than one bid increment above the bid of the advertiser in position (i + 1),**

- **Adopt this principle**
  - An advertiser in position i pays a price per click equal to the bid of an advertiser in position (i+1) plus a minimum increment (typically $0.01).

Conclusion

Many active research areas & open problems

- query understanding
- content matching
- sentiment analysis
- online modeling
- massive optimization
- text summarization
- named entity extraction
- computer-human interaction
- economics of ads

References

- Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords, *Edelman, Ostrovsky and Schwartz*
- From query based Information Retrieval to context driven Information Supply, *Andrei Broder*
- Algorithmic Challenges in Online Advertising, *Deepak Agrawal and Deepayan Chackrabarti*
- Just in time contextual advertising, *Anagnostopoulos et al.*
- Internet Advertising and Optimal Auction Design, *Schwarz*