ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Εξόρυξη γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό

Ενότητα # 6: Web Mining

Διδάσκων: Μιχάλης Βαζιργιάννης

Τμήμα: Προπτυχιακό Πρόγραμμα Σπουδών "Πληροφορικής"

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Οικονομικό Πανεπιστήμιο Αθηνών**» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Οι εικόνες προέρχονται … .

Σκοποί ενότητας

Εισαγωγή και εξοικείωση με τις μεθόδους Web personalization and recommendations (collaborative filtering), Web Advertising.

Περιεχόμενα ενότητας

- Web personalization and recommendations (collaborative filtering)
- Web Advertising

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Web personalization and recommendations (collaborative filtering)

Μάθημα: Εξόρυξη γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό

Ενότητα # 6: Web Mining

Διδάσκων: Μιχάλης Βαζιργιάννης

Τμήμα: Προπτυχιακό Πρόγραμμα Σπουδών "Πληροφορικής"

Web personalization and recommendations

- ~25% of Internet users reading online reviews prior to paying for an offline service,
  - 80% claimed reviews had significant influence on their purchasing habits.
- Users pay a mark-up of 20% to 100% for services/products with excellent peer ratings on review sites.
- Humans are notoriously bad at choosing between too many choices,
  - rely on external recommendations and reviews to narrow the set of possible choices.

Personalization

- Personalized reviews tend to dominate
- Netflix: personalized video-recommendation system based on ratings and reviews by its customers.
- In 2006, offered a $1,000,000 prize to the first developer of a video-recommendation algorithm that could beat its existing algorithm

Recommender Data Model

- Set $U=\{u_1, \ldots, u_n\}$ of users
- Set $I=\{i_1, \ldots, i_m\}$ of items (e.g. products)
- Elements from U and I can be described by a vector respectively
  - $(a_1, \ldots, a_s)$ → attributes of user profile
  - $(b_1, \ldots, b_t)$ → description of items (meta data, features, …)
- Goal of recommendation process: recommend new items for an active user u
- Overview of process
  - User modeling (explicit or implicit, e.g. user rates items)
  - Personalization, generate list of recommended items

User-Item Ranking

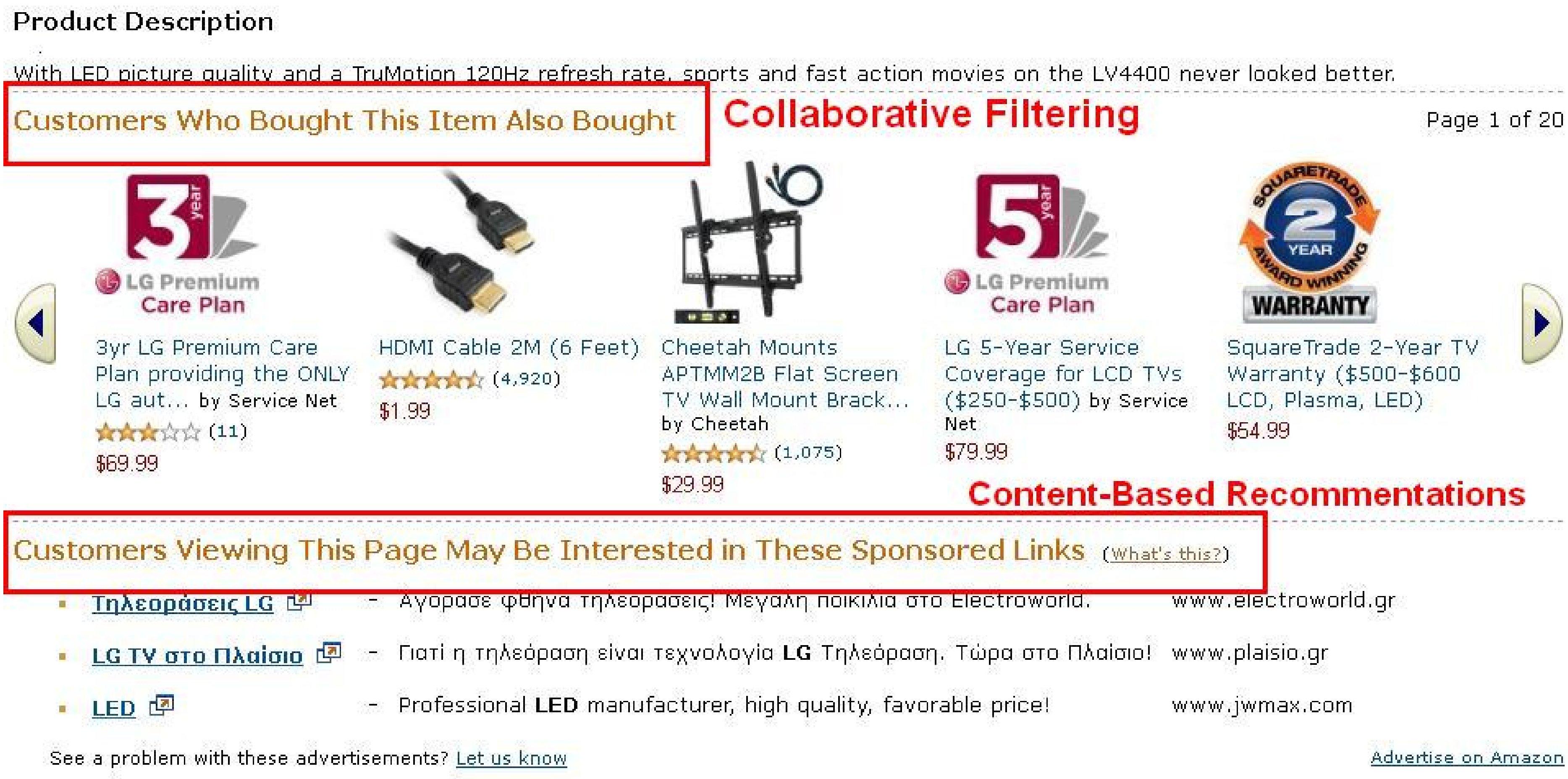
- Recommendation often based on ratings of an item ij by a user u_k:
- Rating $r_{j,k}$: I → [0,1] ∪ ø
- Other range of values possible, e.g. {*, **, ***, ****, *****}
- ø := no rating for Item (or “0”)
- Example user-item matrix of ratings

	V for Vendetta	La Vita e Bella	Lion King	Wall-e
Alice	4	3	2	4
Bob	ø	4	5	5
Cindy	2	2	4	ø
David	3	ø	5	2

Types of Recommender Systems

- Collaborative filtering (CF)
- Content-based filtering (CB)
  - Individual recommender algorithms
  - Also utility- or knowledge-based approaches
- Case-based recommendation
- Hybrid recommender systems
  - Combination of several other recommenders
- Additional important variants
  - Context-aware and multi-dimensional recommenders
  - Decentralized recommender systems
  - Recommending for groups

Example: Product Page on Amazon



Issues of Recommender Systems

- Cold start and latency problems
- Sparseness of user-item matrix
- Diversity of recommendations
- Scalability
- Privacy and trust
- Robustness
- Utilization of domain knowledge
- Changing user interests (dynamics)
- Evaluation of recommender systems

Cold Start Problems

- "New user" and "new item" problem
- Systems cannot recommend items to new users with no profile or no interaction history
- Same for new items
  - Also "latency problem": items need some time until they can be recommended
- Chicken-and-egg problem
  - Users will not use system without good recommendations
  - No incentive to rate items etc.
  - System cannot generate good recommendations
- Possible solutions
  - include explicit user profiling methods to start interaction

Data Sparseness

- Common situation
  - Lots of users and items
  - But only few ratings
  - Sparseness of user-item matrix
  - Recommender algorithms will not work very well
- In addition, new items are continuously added
  - Users should also rate these items
  - Number of ratings has to keep up with new users and items
- Possible solution
  - Include the automatic generation of ratings
  - Implicit user profiling, use of transaction history of users, e.g. click on a video constitutes a positive rating

Diversity of Recommendations

- Focus usually on generating recommendations as “good” as possible
  - But also important: new, unexpected items
  - Do not recommend items that are already known
  - Do not recommend items that are too similar to already known items
    - E.g. user likes “Lord of the Rings 1” → user possibly also likes “Lord of the Rings 2”, but is this really a useful recommendation?
- Possible solutions
  - Use content-based approaches to easier integrate new items in recommendation process
  - Use collaborative filtering to allow “cross-domain” recommendations

Scalability

- Algorithms are based on matching users and items
  - The more items and users, the higher the computational effort to analyze the data
    - Storage/memory and runtime complexity
    - Alternatively, the quality of recommendations suffer
  - Scalability of recommender systems is an issue in practice
- Problem in particular with memory-based approaches
- Possible solutions include
  - Use model-based approach
  - Limit the number of items and/or users
    - E.g. only consider items that received at least k ratings
  - Pre-compute recommendations for users
    - Will reduce runtime

Privacy and Trust

- Collecting and interpreting personal data, e.g. ratings
  - For example, bought items or visited product Web pages on Amazon
  - Control for users?
    - Bought product may have been gift for other person
  - Privacy problem!
- Tradeoff with recommender quality
  - The more information about the user the system is able to collect, the higher the recommendation quality is in general
- Also trust, how can user trust the quality of a recommended item?
- Possible solutions include
  - Consider social relationships (“social recommender”, “Web of Trust”)
  - Let user control their profile information
  - Explanations of recommendations
    - Why was an item recommended?

Robustness

- Quality of (collaborative) recommenders depends on quality of ratings
  - Manipulation by users possible
    - E.g. by automatic registration of a large number of “users” and ratings
  - Also called “shilling”, “profile injection”
  - Attacks in principle
    - “push”: Aim is to push item(s) by inserting a large number of good ratings
    - “nuke”: Same with negative ratings
- Possible solutions include
  - Make registration for service harder, e.g. request and check personal information
  - Detect attacks and remove corresponding users and ratings
  - Adjust algorithms, some algorithms have proven to be more robust

Utilization of Domain Knowledge

- Systems often regard items in isolation
  - No relationships between items
  - No domain knowledge
- Example: searching for (books or other products on) “baseball”
  - Too many hits → restriction to “baseball technique”, or “baseball player”, for example
    - Based on user model and domain ontology
  - Too few hits → broading to “sport”, for example
- Some approaches in current research literature utilize Semantic Web technologies
  - Build and maintain item ontologies
  - Also for users
    - E.g. „GUMO“ (General User Modeling Ontology)

Changing User Interests (Dynamics)

- User model is often relatively static
- But dynamic evolution over user interests
  - Changes over time, older ratings may not be valid any more
- Also the context of recommendations
  - Example: Mobile restaurant guide
    - Restaurant may be too far away from current position (location)
    - Restaurant may be closed today (time)
  - A good rating for a restaurant after a dinner on a weekend may not be relevant for recommending a restaurant for a quick lunch on a workday
- Solutions in research literature include
  - E.g. explicit distinction between short- and long-term interests
  - Context-aware recommender systems

Evaluation of Recommender Systems

- Goal of personalization is to improve the interaction of users with the system
  - May be subjective, hard to evaluate
- General method for recommender systems
  - Let users rate recommended items and compare actual user ratings with predicted rating
  - Most important metrics
    - "precision": probability rate that users did like recommended items
    - "recall": probability rate that preferred items by users are recommended
  - In addition user studies
    - User evaluate system in questionnaire etc.

Collaborative Filtering (CF)

- Basic idea: System recommends items which were preferred by similar users in the past
  - Based on ratings
    - Expressed preferences of the active user
    - And also other users → Collaborative approach
  - Works on user-item matrix
    - Memory-based or model-based
    - No item meta data etc.!
- Assumption: Similar taste in the past implies similar taste in future
- CF is formalization of “word of mouth“ among buddies

General Process

1. Users rate items
2. Find set S of users which have rated similar to the active user u in the past (→ neighborhood)
   - Similarity calculation
   - Select the k nearest users to the active user
3. Generate candidate items for recommendation
   - Items which were rated in neighborhood of u,
   - but were not rated by u yet
4. Predict rating of u for candidate items
   - Select and display n best items

Example (I)

	Hoop Dreams	Star Wars	Pretty Woman	Titanic	Blimp	Rocky XV
Joe	D	A	B	D	?	?
John	A	F	D		F	
Susan	A	A	A	A	A	A
Pat	D	A		C		
Jean	A	C	A	C		A
Ben	F	A				F
Nathan	D		A		A	

Source: http://www.dfki.de/~jameson/ijcai03-tutorial/

Example (II)

	Hoop Dreams	Star Wars	Pretty Woman	Titanic	Blimp	Rocky XV
Joe	D	A	B	D	?	?
John	A	F	D		F	
Susan	A	A	A	A	A	A
Pat	D	A		C		
Jean	A	C	A	C		A
Ben	F	A				F
Nathan	D		A		A	

Example (III)

	Hoop Dreams	Star Wars	Pretty Woman	Titanic	Blimp	Rocky XV
Joe	D	A	B	D	?	?
John	A	F	D		F	
Susan	A	A	A	A	A	A
Pat	D	A		C		
Jean	A	C	A	C		A
Ben	F	A				F
Nathan	D		A		A	

Required Metrics

- Metric for user-user similarity
  - Mean-squared difference
  - Cosine
  - Pearson/Spearman correlation
- Select set S of most similar users (to active user u)
  - Similarity threshold
  - Aggregate neighborhood
  - Center-based
- Metric to predict the rating of u for an item i

Required Metrics

- Metric for user-user similarity
  - Mean-squared difference
  - Cosine similarity
  - Pearson/Spearman correlation
- Select set S of most similar users (to active user u)
  - Similarity threshold
  - Aggregate neighborhood
  - Center-based
- Metric to predict the rating of u for an item i

User-User Similarity

- Item set I
- Users U,V with u[i] denoting rating of item i by user u
  - the rating vector of user u is denoted by $\vec{u}$
  - the vector norm is denoted by $|\vec{u}|$
  - n is the number of items rated by both U and V

- Mean squared difference:
  - Small values show similar users

$$sim_1(U,V) = \frac{(\vec{u} - \vec{v})^2}{n}$$

- Cosine similarity:
  - Large values show similar users

$$sim_2(U,V) = \frac{\vec{u} * \vec{v}}{|\vec{u}| * |\vec{v}|}$$

Pearson/Spearman Correlation

- Average rating is taken into account
  - The vector of average ratings is denoted by $\overline{\overrightarrow{u}}$
- Not suitable for unary ratings
  - Unary: Item is marked (or not)
    - e.g. “Product was purchased“
  - Binary: good/bad, +/- etc.
  - Scalar: Numerical rating (e.g. 1-5) etc.
  - Consider only items which were rated by both users
- Value

$$sim_3(U,V) = \frac{(\overrightarrow{u} - \overline{\overrightarrow{u}}) * (\overrightarrow{v} - \overline{\overrightarrow{v}})}{|(\overrightarrow{u} - \overline{\overrightarrow{u}})| * |(\overrightarrow{v} - \overline{\overrightarrow{v}})|}$$

Example Calculation

User/item	a	b	c	d	e	f
U	5		3		4	
A	1	1		1		
B	1		3	1		
C	5	2	2		5	4
D		3		2		

$Sim_1(U,V)$	$Sim_2(U,V)$	$Sim_3(U,V)$
-	-	-
16	1	0
8	0.76	-1
2/3	0.98	0.833
∞	∞	∞

Required Metrics

- Metric for user-user similarity
  - Mean-squared difference
  - Cosine
  - Pearson/Spearman correlation
- Select set S of most similar users (to active user u)
  - Similarity threshold
  - Aggregate neighborhood
  - Center-based
- Metric to predict the rating of u for an item i

Neighborhood of Similar Users

- Goal: Determine set S of users which are most similar to the active user u
- Center-based
  - S contains k most similar users
    - Problem: maybe some of the users are not really that similar, if k was chosen too large, deviators possible
- Similarity threshold
  - S contains all users with a similarity bigger than a threshold t
    - Problem: maybe too few users in S
- Aggregate neighborhood
  - Follow similarity threshold method first
  - If S is too small (less than k users)
    - Determine "centroid" of set S and add users which are most similar to centroid (→ less deviators than center-based method)

Required Metrics

- Metric for user-user similarity
  - Mean-squared difference
  - Cosine
  - Pearson/Spearman correlation
- Select set S of most similar users (to active user u)
  - Similarity threshold
  - Aggregate neighborhood
  - Center-based
- Metric to predict the rating of u for an item i

CF Recommender (I)

- Given
  - Set S with most similar users to u
  - s[i] rating of a user (from S) from an item i
- Goal: Predict the rating of u for i
- Easiest option: Arithmetic mean

$$r_1(U,i) = \frac{1}{|S|} \sum_{s \in S} s[i]$$

- Problems
  - Similarity of u with members of S is not taken into account
    - Solution: Weighting based on similarity

CF Recommender (II)

- Different users utilize rating scale differently
  - Solution: Consider deviation from average rating (for user)

$$r_3(U,i) = \overline{u} + \frac{1}{\sum_{s \in S} sim(U,s)} \sum_{s \in S} sim(U,s) * (s[i] - \overline{s})$$

- Note
  - Many variations of algorithms in research literature
    - For various application domains, with different properties

Collaborative Filtering

- Amazon and other commercial service use some form of collaborative filtering
  - Exact method usually not published
- Non-commercial example with published algorithms: http://www.movielens.umn.edu
- Exercise ☺
  - Comprehend calculation for introductory example
  - Substitute 1:=A, 2:=B etc.
  - Calculate predicted rating of user “Joe“ for movies “Blimp“ and “Rocky XV”

Advantages Collaborative Filtering

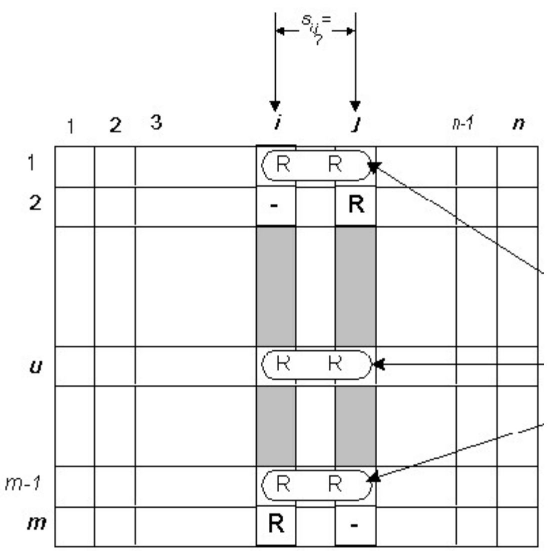
- Works well in practice
- Quality of recommendations improves with density of ratings
- Only ratings as input data required
  - In particular, no information (meta data, description) about items needed
- CF is able to generate cross-domain (“cross genre”) recommendations → high diversity
  - Because item categories etc. are not considered
  - Has proven useful in practice
- Implicit user feedback often adequate (CTR)
  - Unary ratings, e.g. rating = “Click on product Web page”

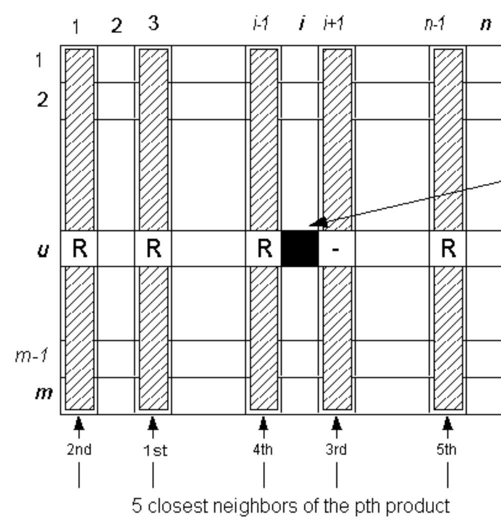
Disadvantages Collaborative Filtering

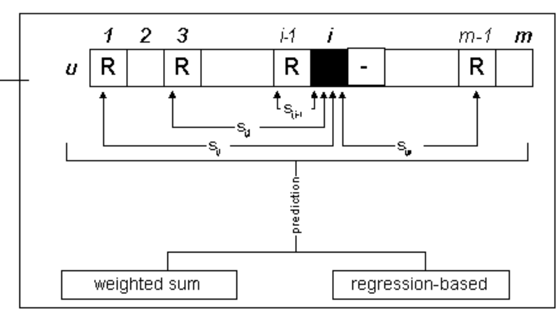
- New user and new item problem
  - Serious issue in practice
- Often sparseness in user-item matrix
  - Algorithms generate worse results with too few ratings
- “Grey sheep” problem
  - Does not work very well for users with “extraordinary” taste
    - Because similar users are not available
  - Also “black sheep”, users that intentionally make incorrect ratings
    - CF is prone to manipulation
    - Trust and robustness are issues

Item-to-Item Collaborative filtering (Amazon)

- Item representation through a N-dimensional vector.
  - Each dimension corresponds to a user's action on this item.
- Rather than matching the user to similar customers, build a similar-items table by finding that customers tend to purchase together.
- Recommend items with high-ranking based on similarity







References

- presentation inspired by slides of Wolfgang Wörndl for “User Modeling, Personalization and Recommender Systems” course in Technical University Munich

  http://www11.in.tum.de/Veranstaltungen/CSCW2:UserModeling,PersonalizationandRecommenderSystems%28IN2119%29

- D. Billsus and M. J. Pazzani, “Learning collaborative information filters”, In Proceedings of the Fifteenth International Conference on Machine Learning, pages 46{54, 1998
- “A Comparative Study of Collaborative Filtering Algorithms”, Joonseok Lee, Mingxuan Sun, Guy Lebanon, http://arxiv.org/pdf/1205.3193.pdf
- A. Paterek. Improving regularized singular value decomposition for collaborative filtering, Statistics, 2007:2{5, 2007.

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Web Advertising

Μάθημα: Εξόρυξη γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό

Ενότητα # 6: Web Mining

Διδάσκων: Μιχάλης Βαζιργιάννης

Τμήμα: Προπτυχιακό Πρόγραμμα Σπουδών "Πληροφορικής"

Advertising

- Why is the advertising important?

  “Advertising is a form of communication that typically attempts to persuade potential customers to purchase or to consume more of a particular brand of product or service. ”

  ---- Wikipedia

The advertising market

- According to <<The Economics>>, the global advertising industry was worth $428 billion in revenues in 2006.
- The global advertising market grew to just over $600 billion in 2007, according to The Kelsey Group.
- The United States is the world's largest advertising market who worth $172 billion in 2008, increased by 53% in last ten years.
- The world's second largest advertising market is China who worth $50 billion, increased by 1200% in the last ten years.
- Followed by Japan who worth $34 billion, UK and German.

Categories of the advertising

- **The traditional one:**

  Based on the traditional media: television, radio, newspapers, billboard.

- **The new one:**

  Based on the internet: Web (online) advertising.

Traditional advertising

Forms:

Television, Radio, Newspaper, Magazine, Billboard, Outdoor, etc.

Traditional advertising

Advantages:

- Huge coverage
- Big spread range

  Example: there are more than 1 billion audiences watched the Beijing Olympic Games Opening Ceremony all over the world!

Defects:

- High investment

  The cost of the advertisement in the Opening Ceremony is about $49,000 per second!

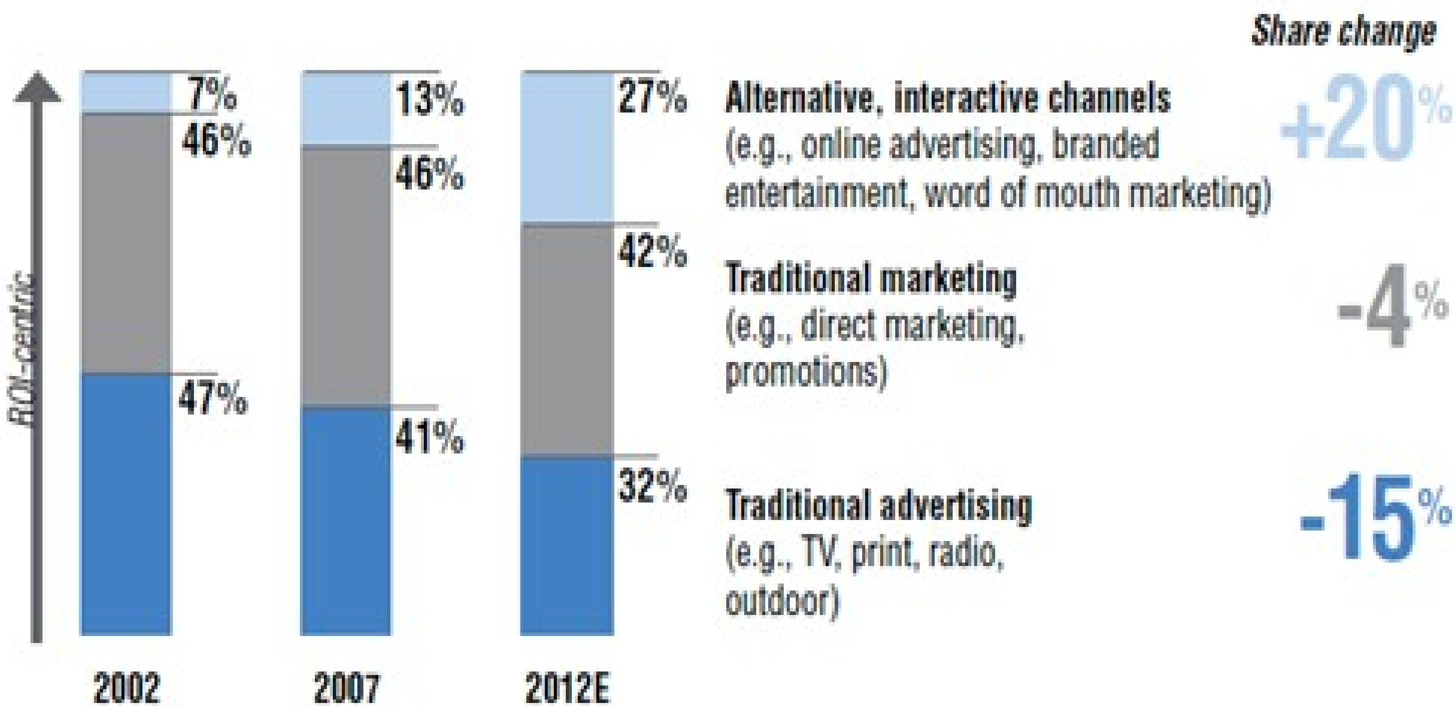
- The ROI (return on investment) is low

  “Half the money I spend on advertising is wasted, the trouble is, I don’t know which half.”

  ---- John Nelson Wanamaker

The traditional advertising is still a major component of the advertising market, however, it is challenged by Online advertising…

FIGURE 2.
U.S. advertising and marketing share (2002 – 2012E).



Source: Veronis Suhler 2008 Communications Forecast; IBM analysis.

Online advertising

- Forms:

  Online advertising is a form of promotion that uses the Internet and World Wide Web for the expressed purpose of delivering marketing messages to attract customers.

  ------ Wikipedia

- Categories:
  - CPI
  - CPC
  - CPA

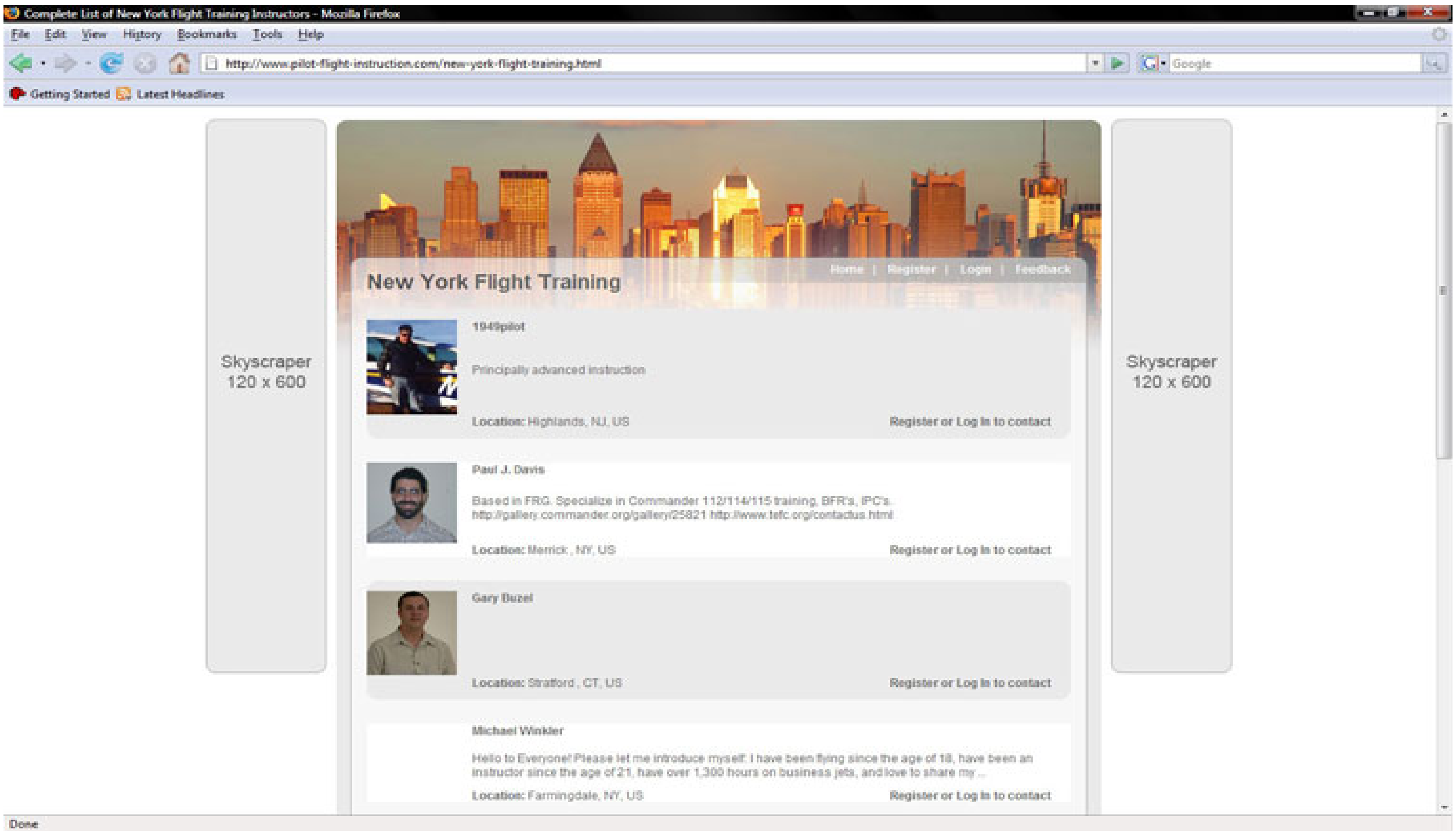
Online advertising

- **CPI (CPM)**

  Cost Per Impression, often abbreviated to CPI, is a phrase often used in online advertising and marketing related to web traffic. It is used for measuring the worth and cost of a specific e-marketing campaign. It is also called CPM, Cost Per Mille. “Per mille" means per thousand impressions.

Online advertising

- Example

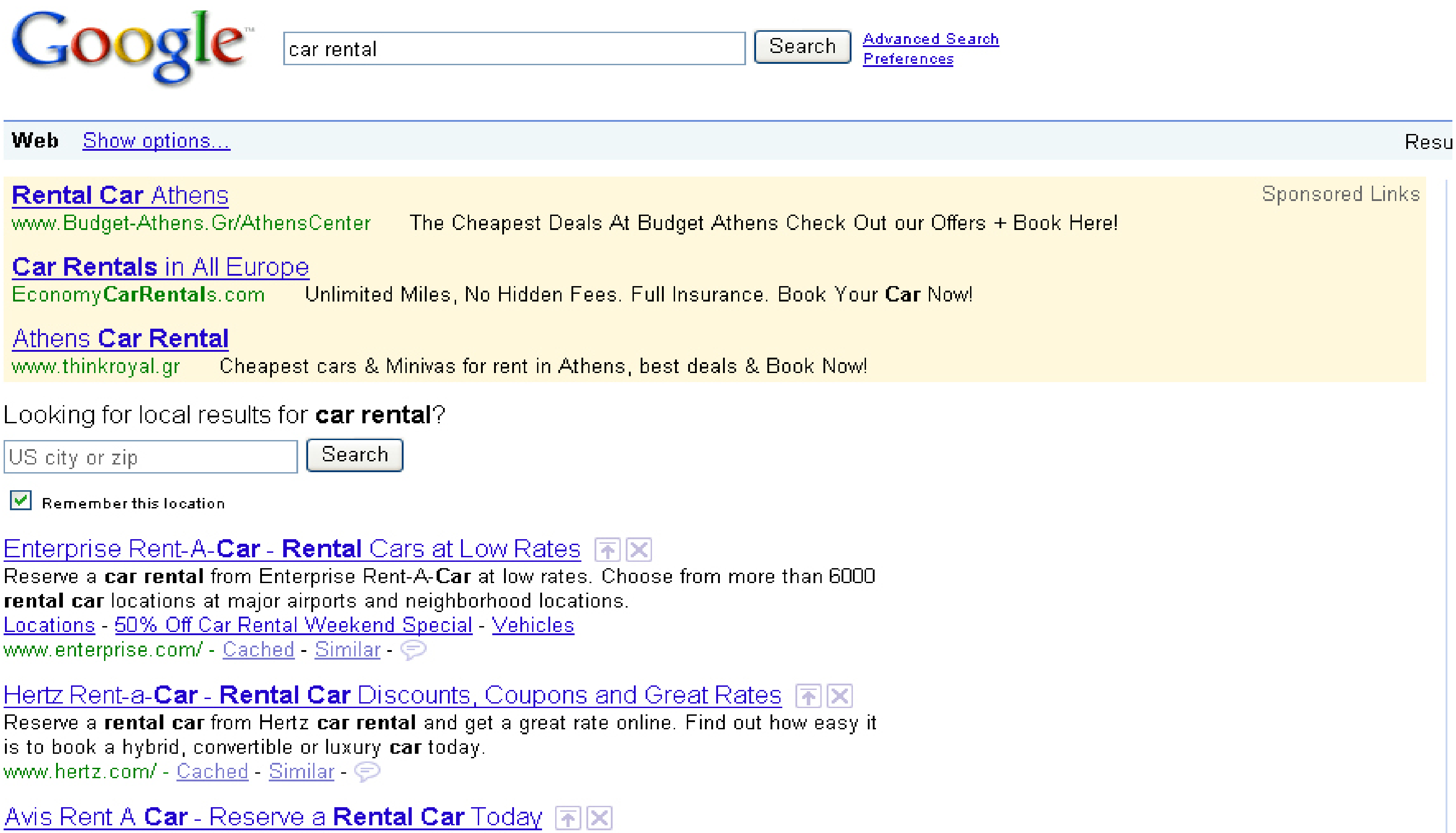


Online advertising

- **CPC**

  Cost Per Click (CPC) is the amount an advertiser pays search engines and other Internet publishers for a single click on its advertisement that brings one visitor to its website.

Online advertising



Online advertising

- **CPA**

  Cost Per Action or CPA (sometimes known as Pay Per Action or PPA) is an online advertising pricing model, where the advertiser pays for each specified action (a purchase, a form submission, and so on) linked to the advertisement.

Online advertising

- Online advertising is targeted.

  Ensure that ad viewers are the ones most likely to buy.

- Online advertising enables *good conversion tracking.*

  Tracking the reach of newspaper and television advertisements is difficult. However, internet advertising allows advertiser to track:

- number of impressions (how many people see it),
- # visits their business web site gets from particular ads,
- conversion rates internet advertisements are getting.

y.

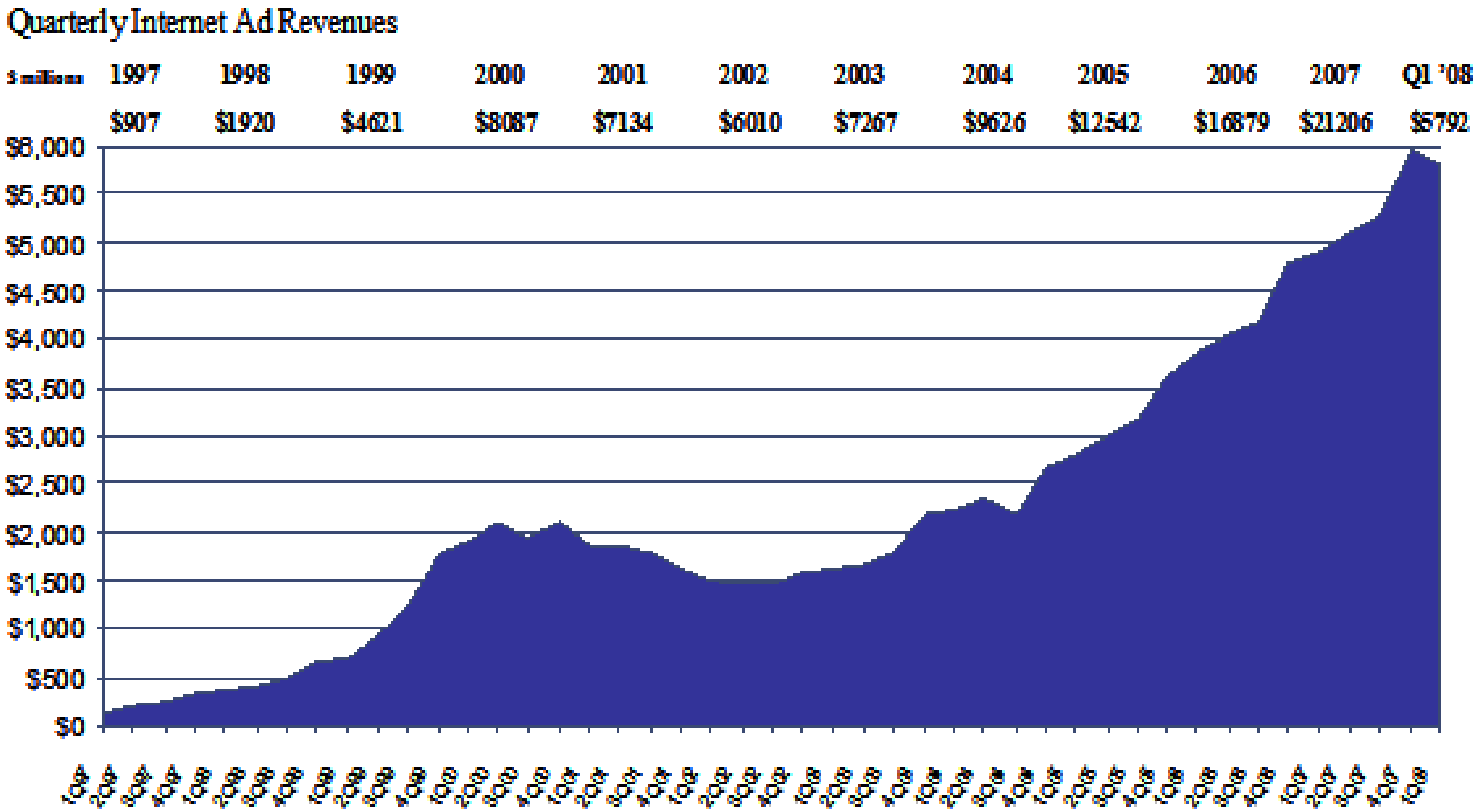
Online advertising

- Online advertising can be much cheaper.

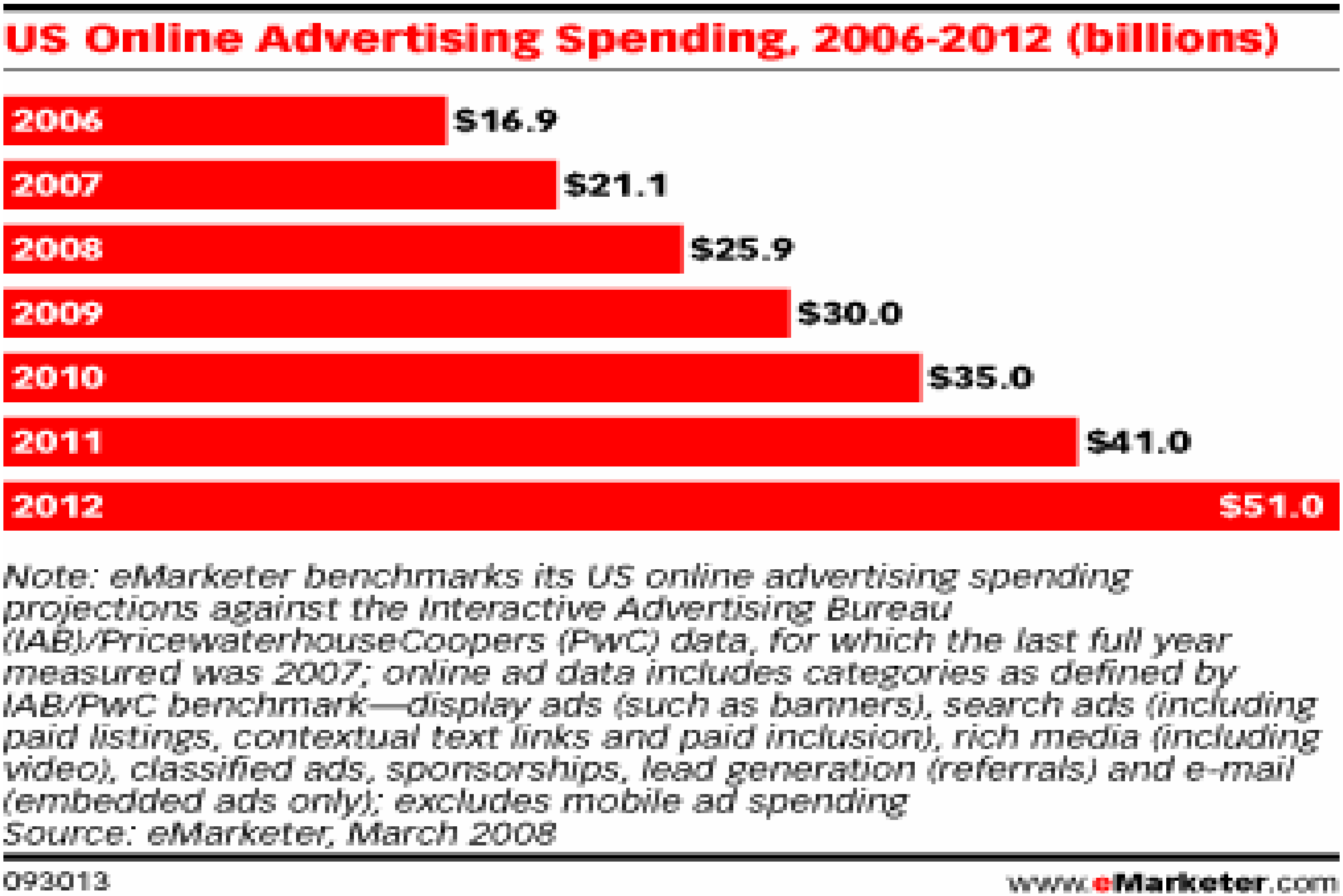
  Because of the targeted nature of internet advertising and the ability to track the effectiveness of ads, conversion rates from internet advertising is typically much better than traditional mediums.

  So the ROI can be much higher.

Online advertising market



Online advertising market



Conclusion

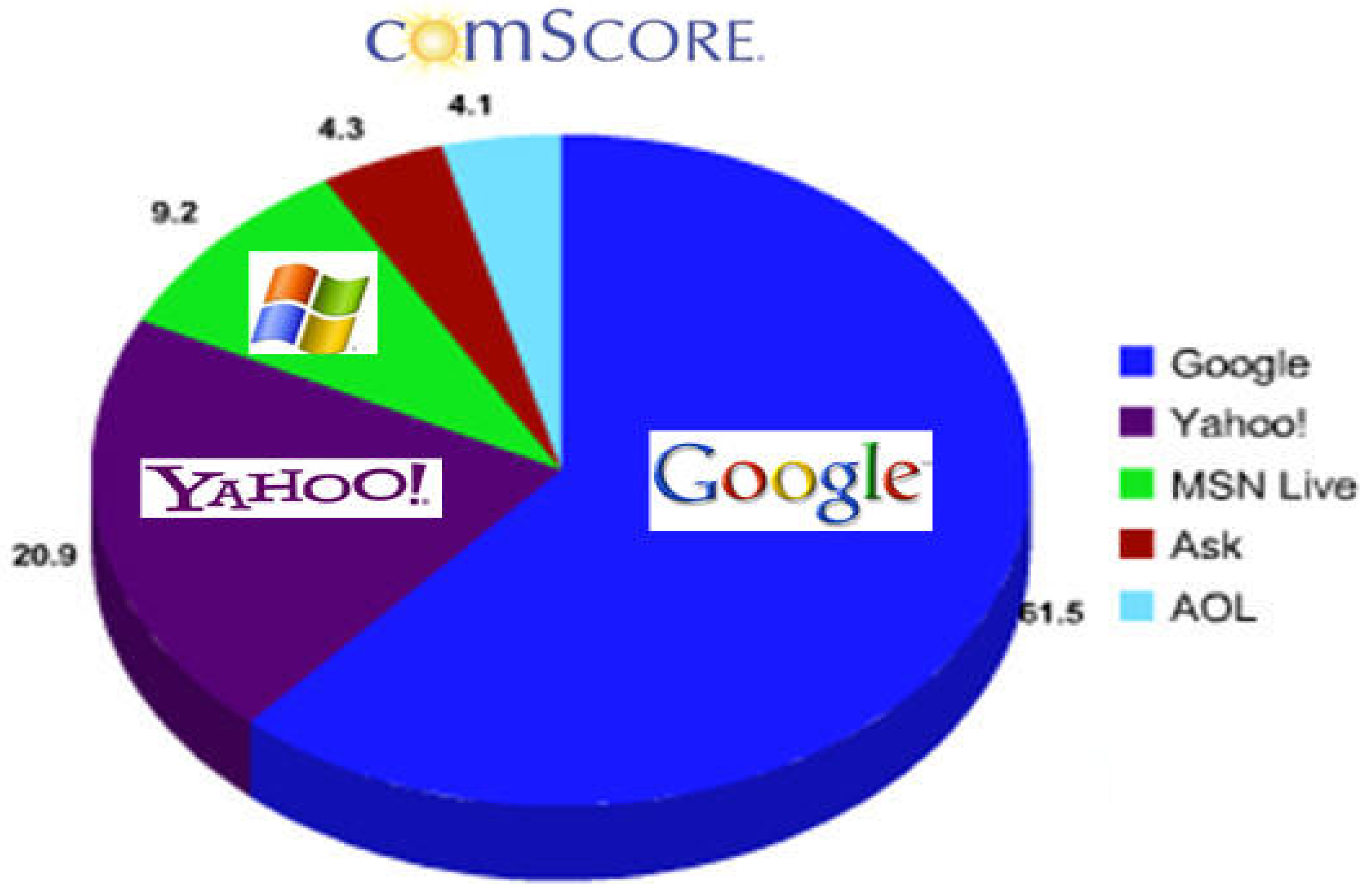
- Online advertising spreads fast.

  its efficiency is much higher than the traditional way.

- Online advertising can track advertising effectiveness.

- Online advertising has a high ROI (return on investment)

Search engine market share



2008

Search engine market share

- The search engine giant ---- Google.
- Google is the most widely used search engine on the internet today. More than 60% of internet searches done online is via Google in the world. It’s market share has increased by 15% in the last 3 years!
- In UK, Google has gained 79% of the search engine market!
- In USA, Google maintained 72% of the search engine market, increased by nearly 30% since last three years!

Web Advertising

- Google ---- the most powerful search engine in the world.
- More than 60% of internet searches done online is via Google in the world. Which means, there are more than 200 million queries searched on Google everyday!
- Google is a platform which collect a great popularity, based on this, it's an ideal intermediate for the dissemination of information, included the advertisements.

Web Advertising

- The three most common ways of web advertising:
  - Cost Per Impression (CPI)
  - Cost Per Action / Acquisition (CPA)
  - Cost Per Click (CPC) / Pay Per Click (PPC)

Google AdWords

What is Google AdWords?

---- Google's flagship advertising product.

- In 2003 Google introduced site-targeted advertising ---- Google AdWords.
- AdWords offers CPC advertising, and site-targeted advertising for both text and banner ads.
- AdWords also offers CPI advertising.
- This advertising product became the main source revenue of Google, which brought a revenue of $50.6 BN$ in 2013

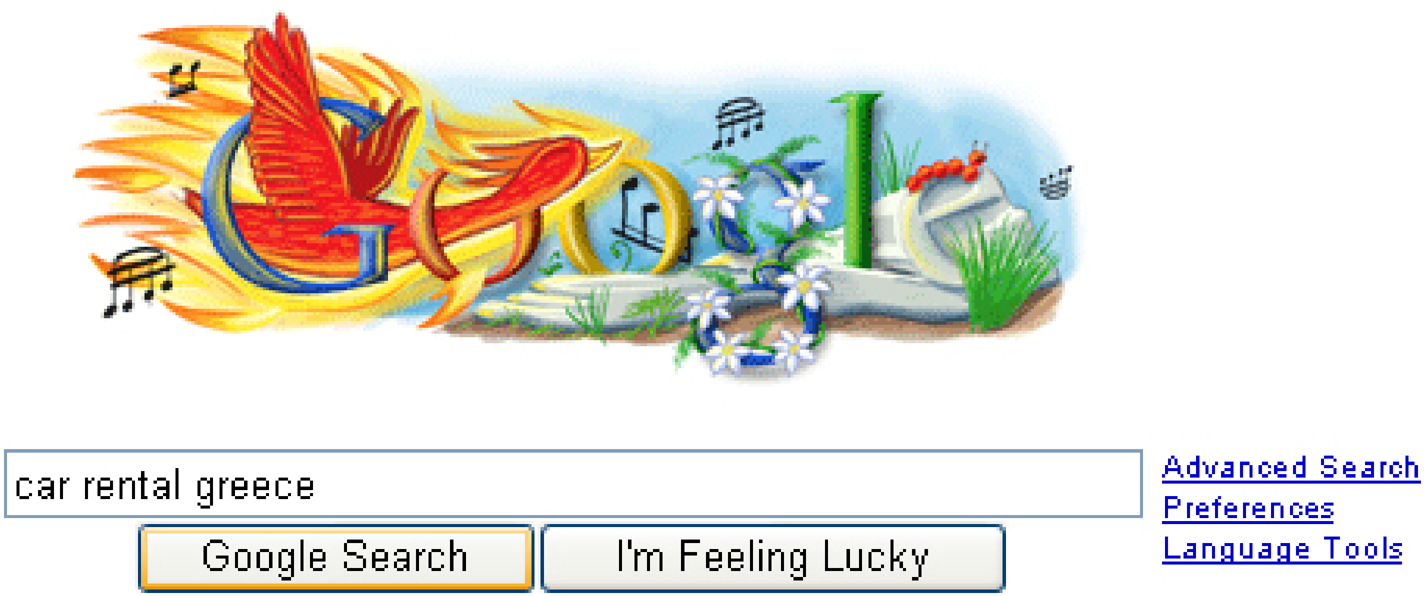
Google AdWords

How does it work?

- Using the AdWords control panel, advertisers can enter keywords, domain names, topics, and demographic targeting preferences, and Google places the ads on what they see as relevant sites within their content network.
- If domain names are targeted, Google also provides a list of related sites for placement.
- Once the somebody searches a keyword on Google, besides the natural results, Google will display the relevant advertisements on the other side.
- Example:

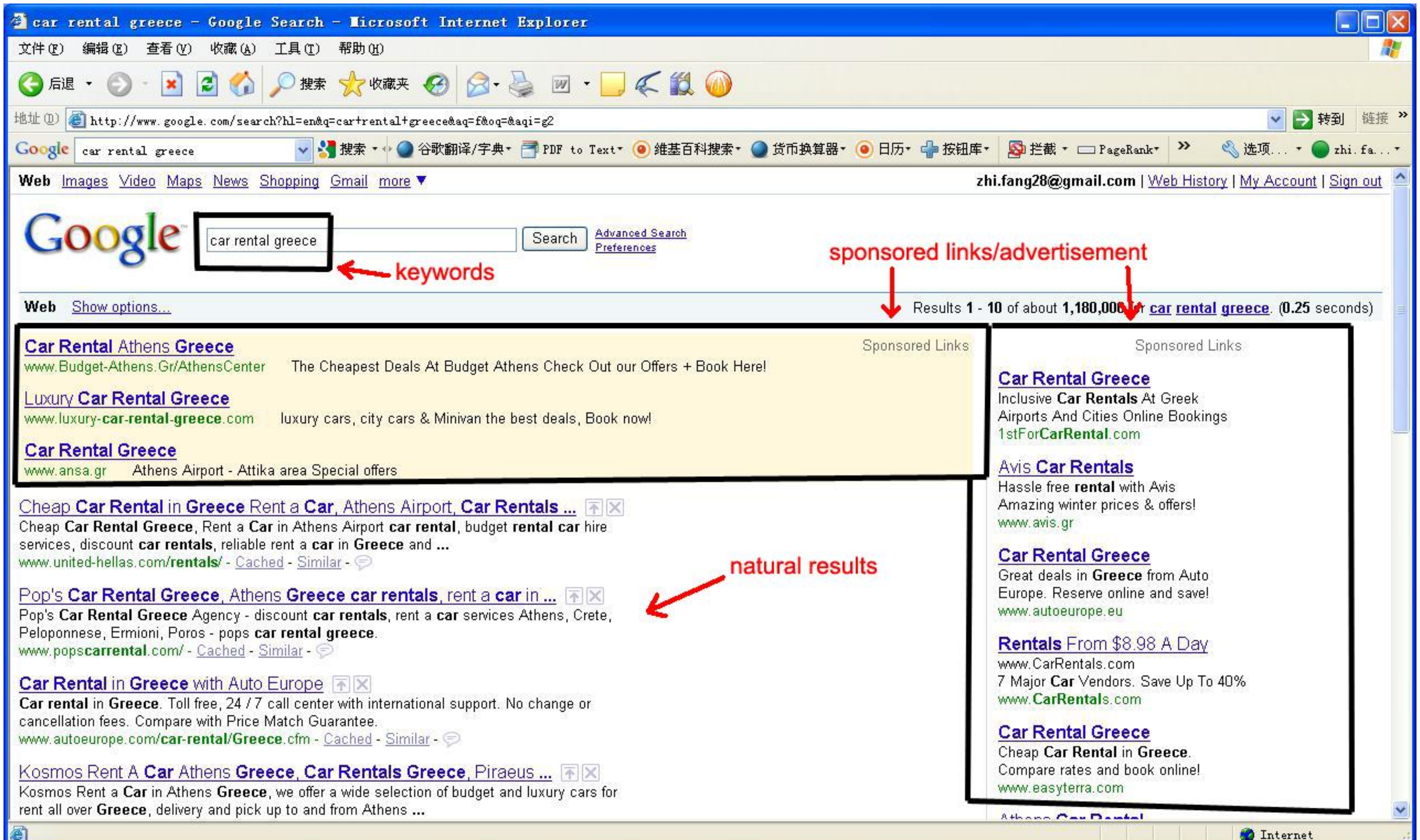
Google AdWords

1. When somebody searches on Google for a particular product or service…



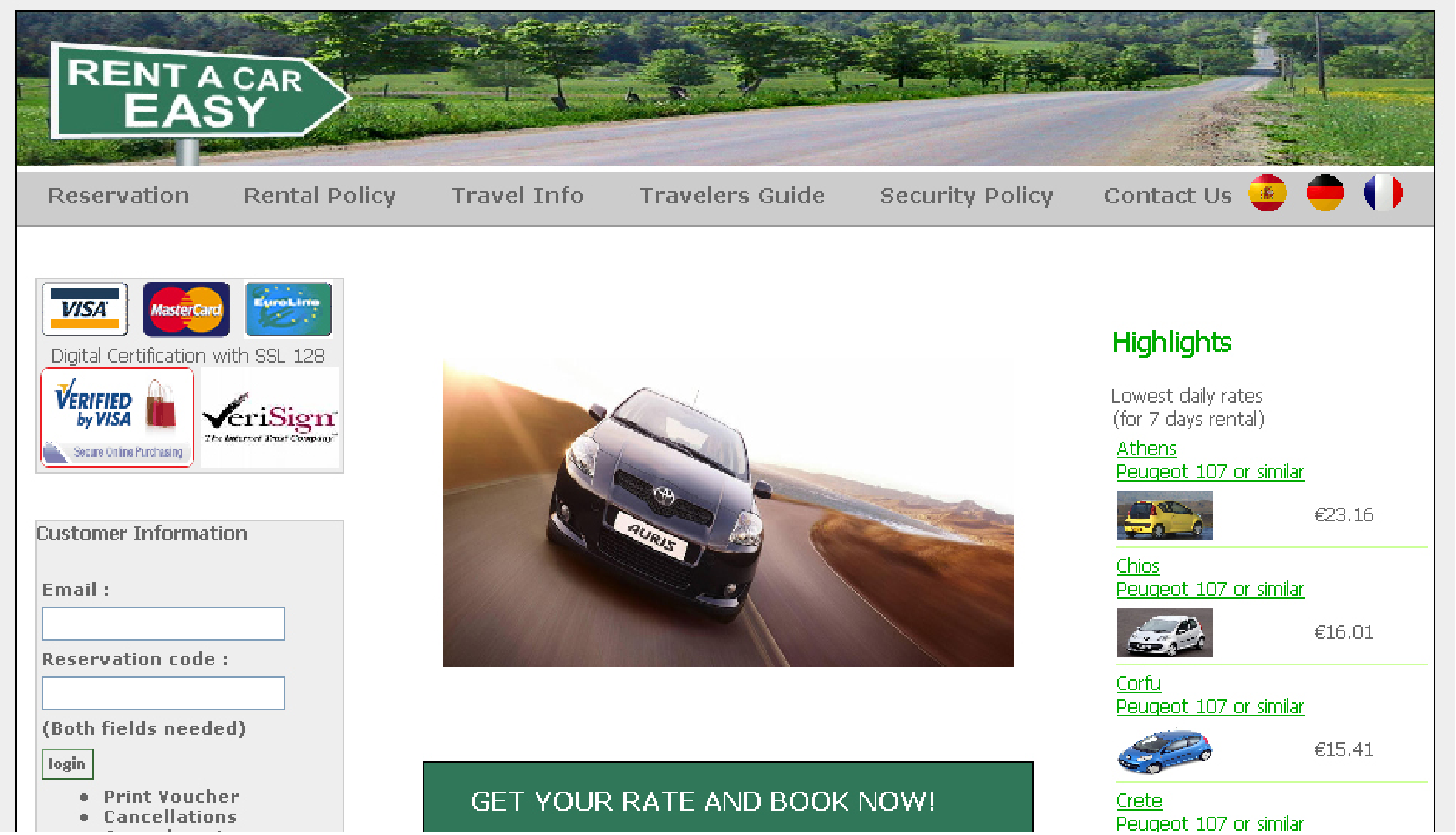
Google AdWords

2. The results given by Google…



Google AdWords

3. Once a clicks on advertisement…



Google AdWords

What are the benefits of using Google AdWords?

- High popularity, huge number of potential customer.

  Google is the most powerful search engine in the world, more than 60% search engine market share.

- High ROI.

  Search engines drive extremely targeted traffic. He who finds your site through a search engine is already actively looking for exactly what you provide.

Google AdWords

What are the benefits of using Google AdWords?

- Board range, variety in forms, easy to implement.

  The AdWords program includes local, national, and international distribution. Google's text advertisements are short, consisting of one title line and two content text lines. Image ads can be one of several different Interactive Advertising Bureau (IAB) standard sizes.

- Advertisers also have the option of enabling their ads to show on *Google's partner networks*. The "search network" includes AOL search, Ask.com, youtube.com, etc.

Google AdWords

How to use Google AdWords?

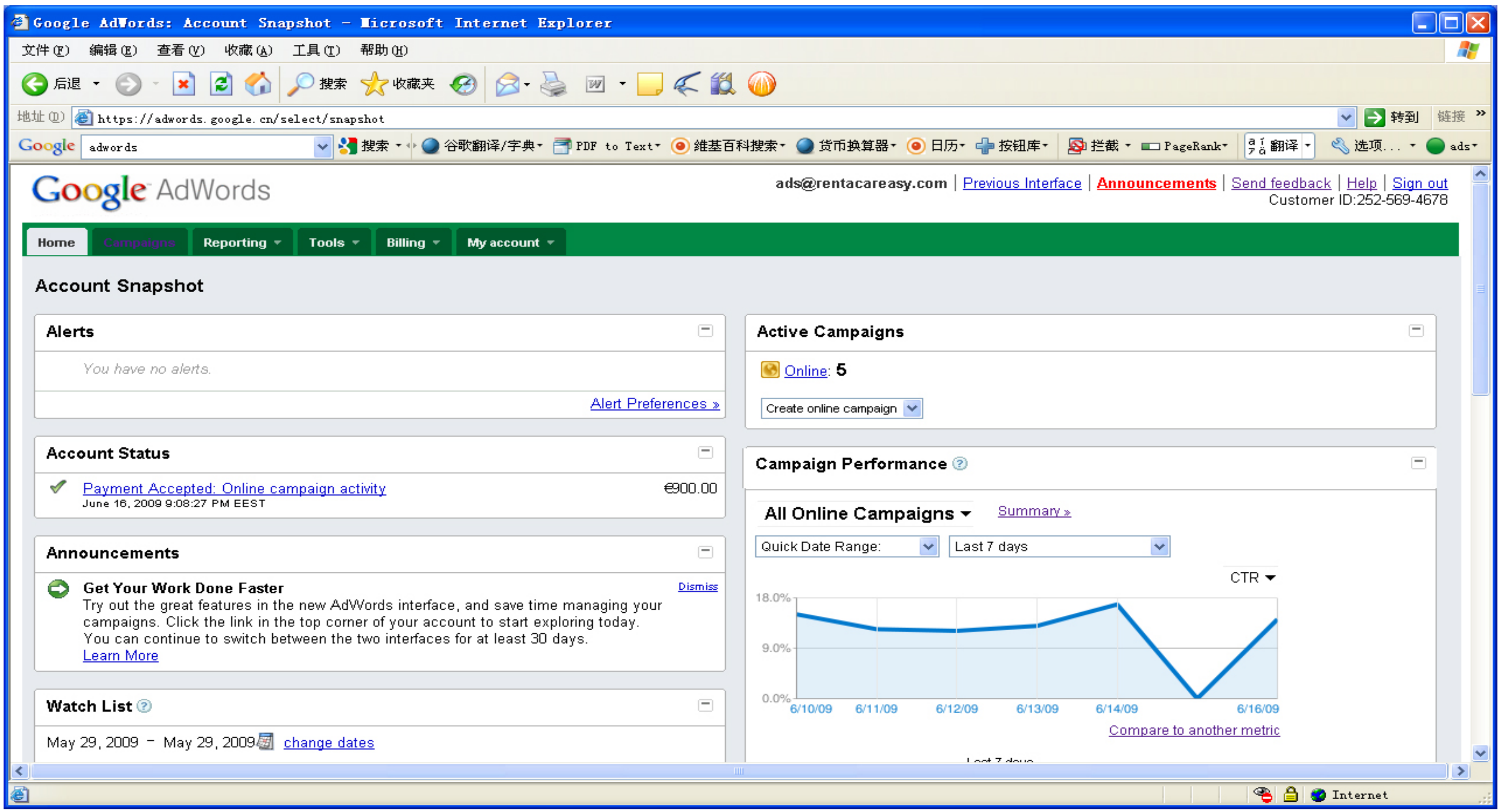
1. Create your own account.
2. Getting start with Organization, Keywords, Placements and Ad Text.
3. Set the maximum CPC bid ---- the bid cost.
4. Improve your quality ---- quality score.
5. Improve the rank of your ad ---- ad rank.
6. Pay the actual cost.

Google AdWords

	Starter Edition	Standard Edition
Simplified sign-up process Sign up with a minimized one-page form.	X	
One product or service Advertise a single product or service with one set of keywords and one or more ads.	X	
Many products or services Create campaigns for multiple products or services, each with many sets of keywords and ads.		X
Multiple ad formats Create text ads, image ads, and other rich ad formats.		X
Basic reporting See a one-page overview of the impressions, clicks, and costs for your ads.	X	
Advanced reporting See a complete library of reports for all aspects of your account. Create custom reports to analyze your costs and return on investment.		X
Basic targeting Target customers in one specific region (like a single country or city).	X	
Advanced targeting Target customers in many regions at once.		X
Advanced cost control Choose from many bidding options: keyword-specific bidding, content bidding, ad position preference, and more.		X
Advanced planning tools Boost your campaign performance with advanced features like conversion tracking, the AdWords traffic estimator, and helpful variations and statistics from the Keyword Tool.		X
Placement targeting Place your ads on the specific websites that appeal to your customers.		X

Google AdWords

2. Getting start with Organization, Keywords, Placements and Ad Text.



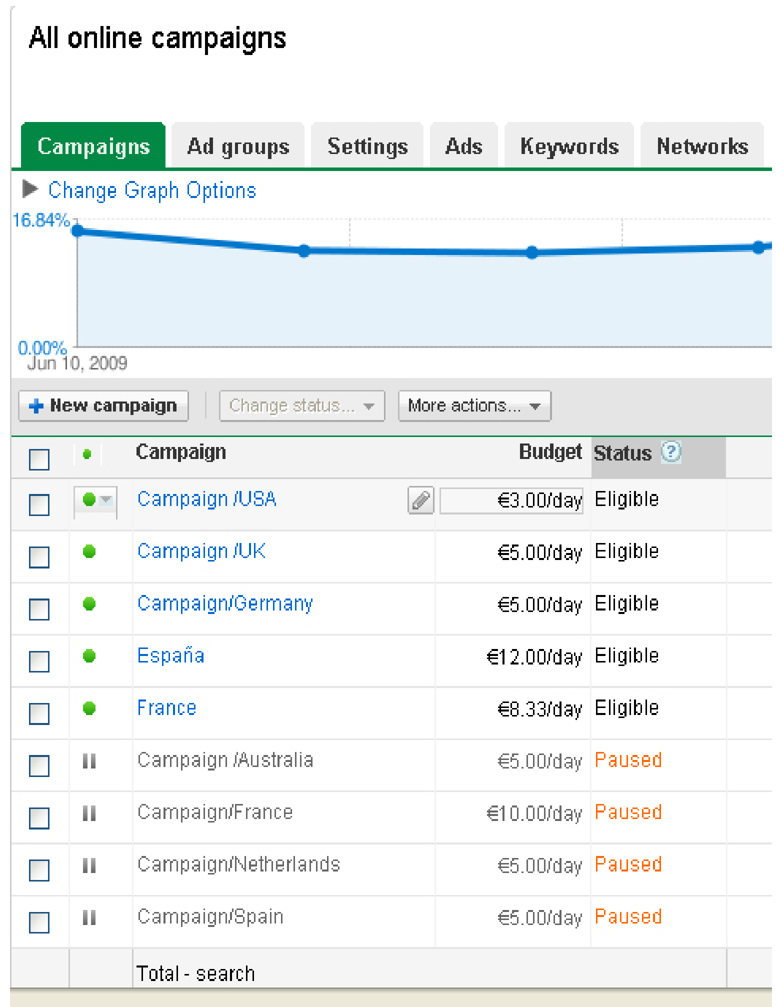
Google AdWords

2. Getting start with Organization, Keywords, Placements and Ad Text.

Campaign Strategy

Every account starts with a single campaign. Each campaign — whether you have one or multiple — should reflect a single, goal.

- target a certain audience,
- sell more products,
- increase signups,

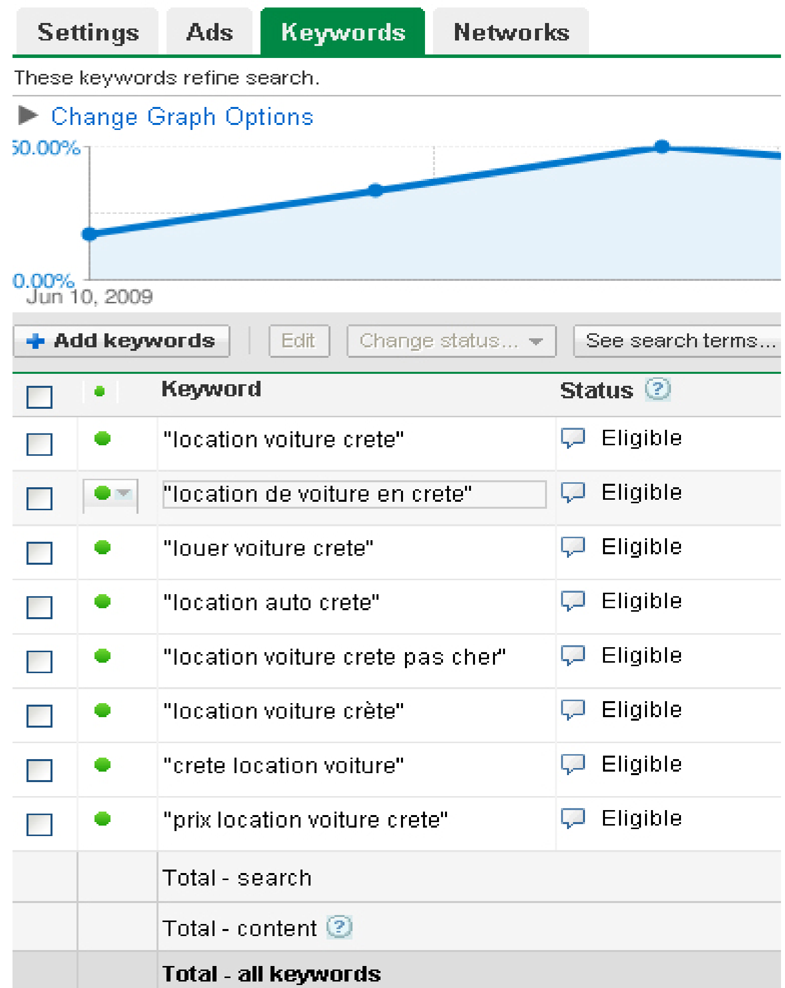


Google AdWords

2. Getting start with Organization, Keywords, Placements and Ad Text.

Ad Group Strategy

Just like your campaigns, your ad groups should be organized by common theme, product, or goal. Often, picking keywords and placements can lay the groundwork for your ad group strategy.



Google AdWords

2. Getting start with Organization, Keywords, Placements and Ad Text.

Ad Text

Google AdWords

3. The bid cost ---- you usually pay less than this amount.
   - With Google AdWords, you set a cost-per-click (CPC) bid or cost-per-1000-impressions (CPM) bid. However, the AdWords Discounter works so you usually end up paying less than this amount.
   - AdWords Discounter calculates actual CPC or CPM. This is the actual amount you pay to maintain your ad's position above the next lower ad. Your actual CPC or CPM is never more than the maximum CPC or CPM bid you specify.

Google AdWords

3. The bid cost ---- Maximum CPC

Your maximum cost-per-click (CPC) is the highest amount that you are willing to pay for a click on your ad. You can set a maximum CPC at the keyword- or ad group-level. The AdWords Discounter automatically reduces this amount so that the actual CPC you are charged is just one cent more than the minimum necessary to keep your position on the page.

	Keyword	Status	Max. CPC
	"location voiture crete"	Eligible	auto: €1.30
	"location de voiture en crete"	Eligible	auto: €1.30
	"louer voiture crete"	Eligible	auto: €1.30
	"location auto crete"	Eligible	auto: €1.30
	"location voiture crete pas cher"	Eligible	auto: €1.30
	"location voiture crète"	Eligible	auto: €1.30
	"crete location voiture"	Eligible	auto: €1.30
	"prix location voiture crete"	Eligible	auto: €1.30
	Total - search		
	Total - content		
	Total - all keywords		

Google AdWords

4. Quality Score ---- the higher, the better

   The AdWords system calculates a 'Quality Score' for each of your keywords. It looks at a variety of factors to measure how relevant your keyword is to your ad text and to a user's search query. A keyword's Quality Score updates frequently and is closely related to its performance. In general, a high Quality Score means that your keyword will trigger ads in a higher position and at a lower cost-per-click (CPC).

Google AdWords

4. Quality Score ---- the higher, the better
   – A Quality Score is calculated every time your keyword matches a search query -- that is, every time your keyword has the potential to trigger an ad.
   – If the campaign uses cost-per-thousand-impression (CPM) bidding, Quality Score is based on:

     The quality of your landing page
   – If the campaign uses cost-per-click (CPC) bidding, Quality Score is based on:

     The historical CTR of the ad on this and similar sites

     The quality of your landing page
   – The best way to improve your keywords' Quality Scores is by optimizing your account.

Google AdWords

5. Ad Rank.

   Ads are positioned on search and content pages based on their Ad Rank. The ad with the highest Ad Rank appears in the first position, and so on down the page.

   Up to three AdWords ads are eligible to appear above the search results (as opposed to on the side). Only ads that exceed a certain Quality Score and CPC bid threshold may appear in these positions. If the three highest-ranked ads all surpass these thresholds, then they'll appear in order above the search results. If one or more of these ads don't meet the thresholds, then the next highest-ranked ad that does will be allowed to show above the search results.

Google AdWords

5. Ad Rank.

Ad Rank formulas

A keyword-targeted ad is ranked on a search result page based on the matched keyword's maximun CPC bid and Quality Score.

Ad Rank = CPC bid × Quality Score

Google AdWords

5. Ad Rank.

Improving your ranking

- Having relevant keywords and ad text,

- a strong CTR on Google,

- a high CPC bid will result in a higher position for your ad.

Because this ranking system rewards well-targeted, you can't be locked out of the top position as you would be in a ranking system based solely on price.

- AdWords Discounter monitors competition and automatically reduces actual CPC so you pay the lowest price possible for your ad's position on the page.

Google AdWords

6. The actual cost

   - never pay more for a click on your ad than the matched keyword's maximum CPC bid (for search pages) or the ad group's content bid (for content pages).

   - quality-based pricing system ensures that you'll often pay less than that amount.

Google AdWords

6. The actual cost

Formula

For search pages, Ad Rank is calculated by multiplying the matched keyword's CPC bid by its Quality Score. For content pages, Ad Rank is calculated by multiplying the ad group's content bid by its Quality Score.

Actual CPC = (Ad Rank to beat / Quality Score) + $0.01

Google AdWords

- Example:
  - Assuming you bid $4/CPC for the keyword “car rental Greece”, with a quality score of 5.
  - And your competitor bids $5/CPC with his quality score equals 3.
  - So your pagerank will be:

    4*5=20

    And your competitor’s will be:

    3*5=15

Google AdWords

– As you have a higher pagerank, your ad will be displayed in front of your competitor's.

– But the actual CPC is:

  15 (the pagerank of advertiser behind you) / 5 (your quality score) + 0.01euros = $3.01

– So the actual price you pay for each click is lower than your bid!

Τέλος Ενότητας # 6

Μάθημα: Εξόρυξη γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό, **Ενότητα # 6:** Web Mining

Διδάσκων: Μιχάλης Βαζιργιάννης**, Τμήμα:** Προπτυχιακό Πρόγραμμα Σπουδών "Πληροφορικής"

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης