Automated Creation and Optimization of Online Advertising Campaigns

Stamatina Thomaidou

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Ph.D. Thesis Defense
Advisor: Prof. M. Vazirgiannis

October 30, 2014

Motivation

- Online Advertising
  - A new and more targeted mean for promotion
  - Different types and channels
  - **Sponsored Search Advertising (SSA)**
    - Primary channel
    - The advertisers have to build campaigns for distributing their ads
- SSA Campaigns
  - Inherent competitiveness
  - Laborious tasks

Objective of PhD Thesis

- Propose a framework which can contribute to considerably optimizing the resources devoted to developing and monitoring a campaign
  - Which keywords?
  - Which ads?
  - Which strategy?
  - Can we automate the full lifecycle?

Outline

- Introduction - Preliminary Online Advertising Concepts
- Keyword Generation
- Ad-Text Generation
- Campaign Optimization (Budget Optimization)
- The Adomaton Prototype
- Conclusions and Future Directions

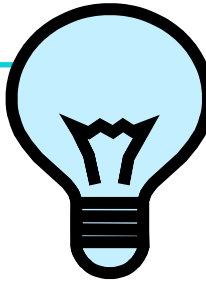
Participants of the Sponsored Search Advertising

Auctioneer (e.g., Google)

- The auctioneer organizes the ad auction process

Advertisers

- They want to promote a product or service
- Each advertiser wants to have better performance than the others



Publishers (in SSA case, equivalent with the **auctioneer**)

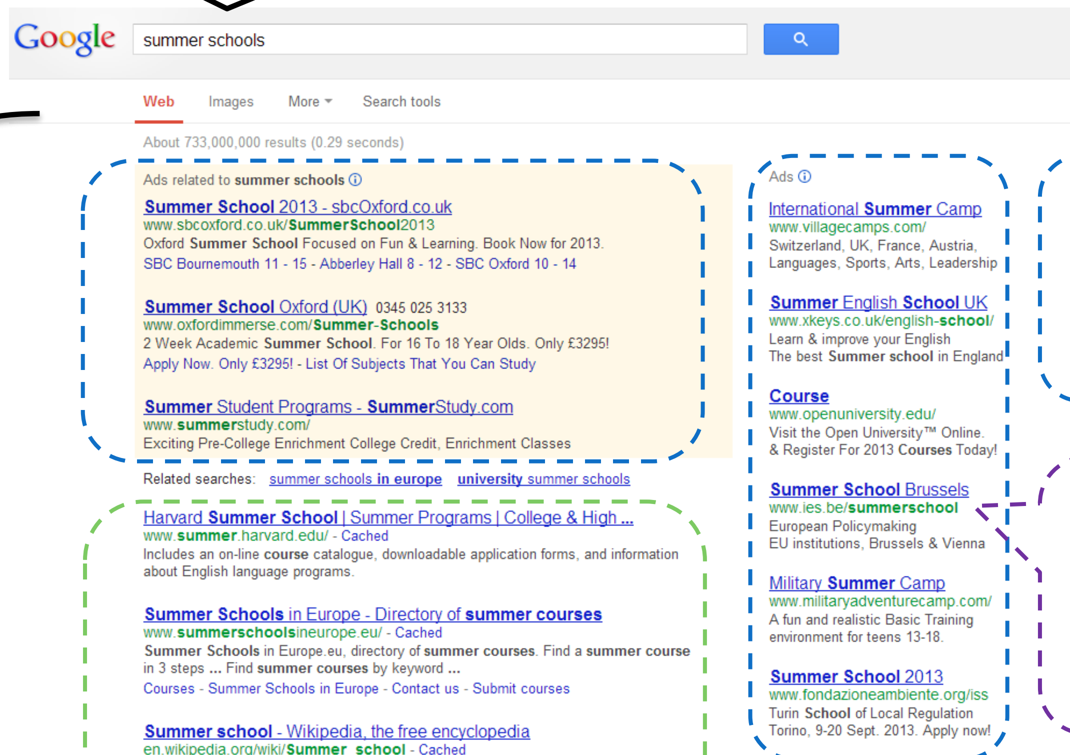
- They give some space for ads (pictures or <u>text</u>)
- In SSA, slots in the search engine results page (SERP)

Users

- The advertisers want to attract their attention
- The users may search on a search engine

Outcome of an Ad Auction (1/3)

Query / Keywords:
initiating the ad auction



Search Engine
Result Page
(SERP)

Sponsored Results
- Ads
- Textual Advertising: Ad-text (ad creative or ad copy)

Landing Page
- Click the ad
- Make an “action” (conversion) inside the page

Algorithmic / Organic Results

Outcome of an Ad Auction (2/3)

- Impression
  - The appearance of an advertisement in a SERP after a user's query
- Click-Through Rate (CTR)
  - The percentage of people clicking on an advertisement when it appears in a SERP
  - CTR = Clicks/Impressions
- Conversion Rate (CR)
  - The percentage of conversions against clicks
  - CR = Conversions/Clicks

Outcome of an Ad Auction (3/3)

- Bid (or maxCPC)
  - The maximum amount of money that an advertiser is willing to pay for a click
- Cost per click (CPC or avgCPC)
  - The actual amount of money that an advertiser is being charged for a click on his advertisement
  - <u>Hybrid second-price auction</u>
- Quality Score
  - Estimate of how relevant to the promoted product are ads, keywords, and landing page

Outline

- Introduction - Preliminary Online Advertising Concepts
- Keyword Generation
- Ad-Text Generation
- Campaign Optimization (Budget Optimization)
- The Adomaton Prototype
- Conclusions and Future Directions

Keyword Generation

- This process aims at proposing valid and representative (advertising) keywords
  - Other definitions in literature and industry
    - Keyword selection
    - Bid phrase generation [Ravi et al. 2010]
    - Keyphrase extraction and suggestion

Keyword Generation Challenges

1. Relevance
   - Quality Score factor
2. Specificity
   - Users who are searching using phrases are more likely to click the ad or convert
   - General keywords are expensive due to high competitiveness
3. Variety of suggestions
   - Extra terms than the standard hand-picked
   - Non-obvious: Might not exist in the landing page but are relevant

Our Contributions

Propose relevant, specific, and nonobvious set of multiword keywords

- *S. Thomaidou, M. Vazirgiannis.* ***Multiword Keyword Recommendation System for Online Advertising.*** *IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM'11), Kaohsiung, Taiwan*

GrammAds Prototype

- *S. Thomaidou, K. Leymonis, M. Vazirgiannis.* ***GrammAds: Keyword and Ad Creative Generator for Online Advertising Campaigns.*** *Digital Enterprise Design & Management Conference (DED&M'13), France, Paris*

Keyword Extraction

- Extract useful information from HTML document (landing page)
  - Start with preprocessing

- Assign **weights** to important **tags**
  - $weight_{tag}$

- Term scoring for term *i*
  - $relevanceScore_i = \frac{\boxed{\sum weight_{tag} * f_{itag}}}{MAX\_WEIGHT}$
  - f_{itag}: frequency of term inside a tag
  - *MAX_WEIGHT:* maximum expected special weight that a term could have

Special weight:
Similar to Okapi BM25F

In need of “Multiword Keywords”

- Advertising keywords
  - While keywords of one term (unigrams) frequently have a broad meaning, phrases (higher order n-grams) are more **specific**
  - A typical query length, especially while searching for a product or service, varies between **1 and 3 terms**

- Build possible combinations of previous extracted unigrams
  - **Co-occurrence**
    - **If term*i* and term*j* appear in a <u>same unit (HTML tag)</u>, then they co-occur once – we increase their f*i,j***
  - Extract top bigrams and trigrams
  - Boost trigrams, then bigrams, then unigrams

Keyword Suggestion

- Discover new non-obvious phrases that are not appearing inside the landing page [Abhishek et al. 2007, Ravi et al. 2010, Joshi et al. 2006]
  - e.g., for a campaign heavily based on “cheap flights” combinations: “low cost airlines”
- For each given seed keyword (extracted from previous step)
  - Keyword is entered as a query into a search engine API (e.g., “car rental”)
  - Result set of short text snippets relevant to the query - Keep most frequent unigrams
  - Score them with a normalized *tf-idf* score
- **Why exploitation of SERP snippets?**
  - **Faster, more compact**
  - **More sure about the semantics**

Comparison with State-of-the-art commercial systems

- The landing pages for our experiments were taken from different thematic areas, promoting several products and services
  - 8 categories – Top-20 results

- Comparison with competitive keyword selection systems
  1. GrammAds
  2. Google Keyword Suggestion Tool
  3. Alchemy API
  4. Google AdWords API RelatedToUrl method

Human Evaluation Guidelines

- Human ranking for resulted keywords following a blind testing protocol
- Criteria
  1. **Relevance** : The relevance of keywords related to each landing page
  2. **Specificity** : How general or specific were the generated keywords
  3. **Nonobviousness** : Overall variety of combinations on the final set;
     How redundant and repeatable or nonobvious and diverse was the final set related to the category and advertising form of each landing page

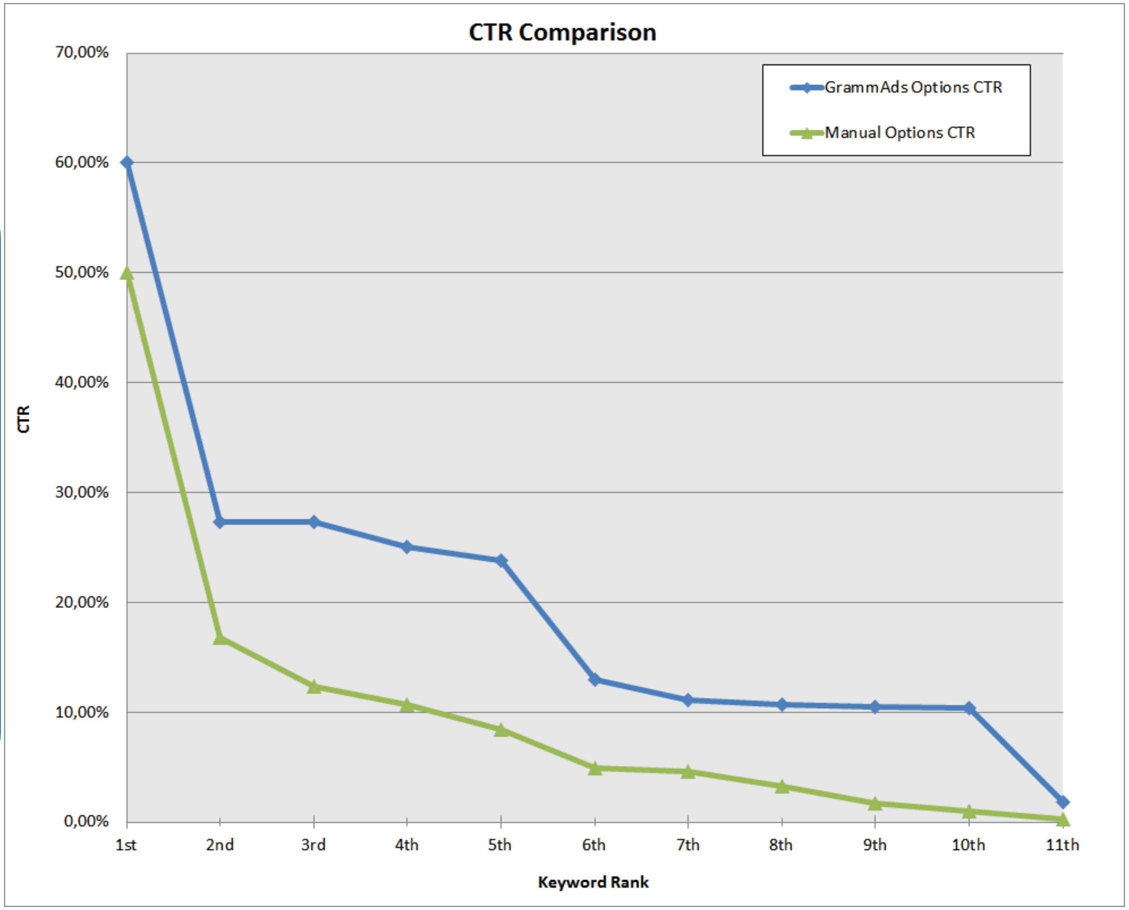
Results

Metric	GrammAds	Google Keyword Suggestion	Alchemy	Adwords RelatedToURL
Relevance	0.76	0.77	0.50	0.33
Specificity	0.61	0.8	0.41	0.21
Nonobviousness	**0.76**	0.67	0.50	0.1
Harmonic Mean	0.70	**0.74**	0.40	0.16

Real-world Campaign Evaluation

Keyword CTR Comparison for top-11 terms

- Manual vs. GrammAds for a period of 2 weeks
- Using the exact same bidding strategy for two identical campaigns for a prefabricated housing company
- ***GrammAds keywords more successful***



Outline

- Introduction - Preliminary Online Advertising Concepts
- Keyword Generation
- Ad-Text Generation
- Campaign Optimization (Budget Optimization)
- The Adomaton Prototype
- Conclusions and Future Directions

Ad-Text Generation

- Concise ad/promotional-text
  - New online advertising models: Sponsored Search, Mobile advertising, Promoted Tweets, etc.
- Automated and massive manner with highest possible quality
  - Campaigns for large websites/e-shops/online catalog aggregators
- Extract content either from a description field or other kind of information feedback
  - Positive reviews/comments

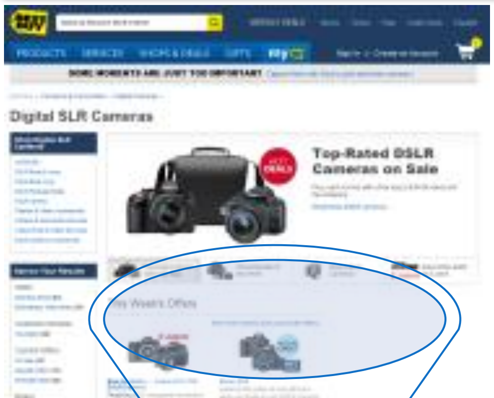
Our Contribution

Produce in an automated manner small comprehensive text ads (promotional text snippets) while maintaining relevance, clarity, and attractiveness

- *S. Thomaidou, I. Lourentzou, P. Katsivelis-Perakis, M. Vazirgiannis.* ***Automated Snippet Generation for Online Advertising.*** *ACM International Conference on Information and Knowledge Management (CIKM'13), San Francisco, USA*

Ad-Text Generation: Proposed Methodology

A landing page relevant to a camera model inside an e-shop



(1) Phrase Extraction

Product name: Fujifilm Finepix JX580
extended battery life, web site constitutes acceptance, instantaneously increases shutter speed, rapidly check recently shot images, additional accidents coverage, artistically enliven photos, electrical failures…

(2) Advertising Language Generation

Constructing a good form of the final ad-text sentences with n-gram form correction, combinations, templates

<optional: product name> <with>

<feature set sequence>

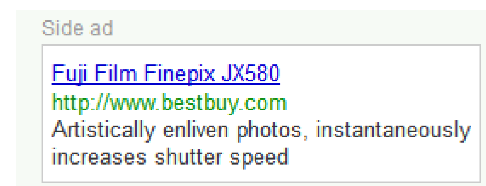
- Artistically enliven photos, instantaneously increases shutter speed
- Artistically enliven photos, extended battery life
- Max Video Resolution 1280, instantaneously increases shutter speed
- (…)
- Max Video Resolution 1280, electrical failures

Advertising Language Model Evaluation

(3) Sentiment Analysis Filtering

Filter out the negative snippets

Top positive snippets



Phrase Extraction

- Preprocessing & retrieve all **unigrams** and their raw frequencies
- **Bigrams**: sequence of two adjacent unigrams
- Merge initial n-grams and formulate gradually higher order n-grams
  - “cheap car” + “car rental” = “cheap car rental”
  - Check Jaccard similarity and keep as a candidate the most **representative** phrase
  - $\sigma_{summaryLength}$ = 70
  - **Make transformations**
    - **Well-formed n-grams**
    - **Templates**

Representativeness Scoring

- How well a phrase represents the given webpage [Ganesan et al. 2012]

$$\text{Representativeness}(w_i \ldots . w_n) = \frac{1}{n} \sum_{i=0, j=1}^{n} pmi'(w_i w_j), i \neq j$$

  - where $w_i \ldots . w_n$: unigrams of the phrase
  - n pairs
  - *pmi'*: modified version of the word-based *pointwise mutual information* [Church et al. 1989]

Measures strength of association between words

$$pmi'(w_i w_j) = \log_2 \frac{p(w_i w_j) c(w_i w_j)}{p(w_i) p(w_j)}$$

Rewards well associated words with high co-occurrence

Advertising Language Generation

- Information Scoring
  - Trade-off between high-scored n-grams and efficient utilization of character space
  - $I(snippet) = \boxed{\frac{1}{n+\frac{\sigma_{summaryLength}}{length(snippet)}}} \sum_{i=1}^{n} \text{Representativeness}(i)$

Ideal case: 1 n-gram exactly 70 chars

- Readability Scoring
  - We constructed a dataset of about 50,000 unique ads obtained from major search engines (queries from Google Products Taxonomy)
  - Language model based on trigrams and on the Kneser-Ney discounting method: logarithmic probability as the value of the Readability Score
- Final Candidates

$$OverallScore(c) = \alpha \cdot I_{norm}(c) + (1-\alpha) \cdot R_{norm}(c)$$

α = 0.5

Sentiment Analysis Filtering

- Inside the landing page might exist also negative reviews or comments that can distort our ads
  - Filter out negative snippets

- Amazon (reviews) Sentiment Dataset Snapshot
  - Balanced dataset
  - Does not contain any neutral reviews (i.e., rated with 3 stars)
  - Each line in the positive and negative set corresponds to a single snippet (usually containing roughly one single sentence)

- Trained a Naïve Bayes Classifier
  - Bag-of-words
  - Train on about 260,000 instances, test on 87,000 instances
  - Accuracy: 0.841

Ad Snippets Evaluation

- Comparison with variations of our principal method and baselines
- 100 Product Landing Pages: 50 from eBay & 50 from BestBuy
- For each landing page we applied 7 methods and we generated 3 text ads (i.e. ad snippets) for each of them
  - Overall, 2100 snippets for evaluation
- 12 Human Evaluators following a blind testing protocol
  - 2 groups
- In each group were given triplets of the form {landing page, product name, ad snippet}

Principal Method Variations

- **PE+ALG:** Phrase Extraction Phase, Complete Advertising Language Phase

- **PE+ALG+SA:** PE+ALG results and elimination of the negative ones

Baselines

- **PE:** Only Phrase Extraction Phase
- **PE+HG:** Eliminating candidates after a minor heuristic grammar check
- **PE+SA:** Eliminating any n-grams that are classified as negative phrases
- **PE+HG+SA:** Eliminating n-grams that are not compatible with both HG and SA checks
- **PE+CP:** Evaluation with Information and Readability Ranking Functions <u>(without transforming n-grams)</u>

Examples of generated promotional text from all methods with positive (good) evaluations for all criteria

Method	Product Name	Snippet
PE	Canon PIXMA iP100	*Auto image fix function automatically adjusts image*
PE+HG	Dell XPS Desktop	*Microsoft windows 8 64-bit operating system preinstalled*
PE+SA	Samsung Galaxy S	*Cell phone free shipping no contract required $25 free extras*
PE+HG+SA	Virgin Mobile - LG Optimus	*Bluetooth compatibility for wireless communication*
PE+CP	Dell Ultrabook	*Corning gorilla glass ensures durability*
PE+ALG	VIZIO ESeries HDTV	***Vizio eseries with effective refresh rate, low price guarantee***
PE+ALG+SA	Fujifilm Finepix JX580	***Artistically enliven photos, instantaneously increases shutter speed***

Human Evaluation Guidelines

- The evaluators were asked to answer if the ad snippet met a criterion (with value 1) or not (with value 0)

- The criteria were the following:
  - Attractiveness *A*: Is the snippet attractive in order to prompt you clicking on the ad?
  - Clarity *C*: Combination of Grammaticality and Readability. Is the snippet structured in a comprehensible form?
  - Relevance *R*: Is the snippet representative for the corresponding landing page of the promoted product?

Interjudge Agreement

Criterion	P(D)	P(E)	kappa statistic
Attractiveness	0.891	0.553	**0.756**
Clarity	0.912	0.561	**0.800**
Relevance	0.944	0.541	**0.877**

- Kappa value around 0.8 is considered as very good agreement

Criteria Rates per Method

- "Relaxed" evaluation: If either one of the judges has assigned a positive value
- H: Harmonic mean for A, C, R rates

Method	Attractiveness	Clarity	Readability	Harmonic Mean
PE	0.387	0.677	0.660	0.538
PE+HG	0.253	0.693	0.517	0.410
PE+SA	0.433	0.697	0.680	0.575
PE+HG+SA	0.293	0.647	0.550	0.443
PE+CP	0.257	0.617	0.423	0.381
PE+ALG	**0.527**	**0.854**	**0.943**	**0.726**
PE+ALG+SA	**0.593**	**0.850**	**0.937**	**0.763**

Results

- Wilcoxon rank-sum test between the most efficient methods and the best performing baselines

- Differences in the A, C, R scores between:
  - *PE* vs. *PE + ALG*
  - *PE* vs. *PE + ALG + SA*
  - *PE + SA* vs. *PE + ALG*
  - *PE + SA* vs. *PE + ALG + SA*
  - **were actually statistically significant ($p < 0.05$)**

- **Remark:**
  - Adding SA does not affect the computational cost
  - It provides an interesting solution for more prompting ads

Outline

- Introduction - Preliminary Online Advertising Concepts
- Keyword Generation
- Ad-Text Generation
- Campaign Optimization (Budget Optimization)
- The Adomaton Prototype
- Conclusions and Future Directions

Budget Optimization for Multiple Keywords

- Find a bidding strategy for the advertiser that maximizes his profit
- The budget of an advertiser needs to be split among several keywords
- Simple or Weighted Keyword Bidding Problem [Feldman et al. EC'07]
- Budget optimization is NP-hard
  - Approximate solution
  - Stochastic models
- Lack of experimentation in real-world campaign data

Our Contribution

We propose an approximate and online/adaptive to the current auction conditions solution on the budget optimization problem for maximizing advertiser's monetary profit or traffic

- *K. Liakopoulos, S. Thomaidou, M. Vazirgiannis.* ***The Adomaton Prototype: Automated Online Advertising Campaign Monitoring and Optimization.*** *Ad Auctions Workshop, ACM Conference on Electronic Commerce (AAW'12-EC'12), Valencia, Spain*

Parameters for the Optimization Model

- The advertising agent has the role of an investor

- The capital is the total budget B for the period that the campaign is active

- The profit from the **clicks** or **conversions** for each investment is represented as v
  1. Value for traffic
  2. Value for monetary profit

- The cost that the advertiser is finally charged for a specific investment is w
  1. Weight for traffic
  2. Weight for monetary profit

Multiple-choice Knapsack Problem Formulation

- Investment : Final item *x* which is a pair (k, b)
  - keyword: k
  - bid: b
- The advertiser has *j* options of (k,b) candidate pairs
  - Only *one pair per investment* for his final proposal
- Total number N of the final chosen investments = r available keywords of the campaign

$$\text{maximize} \quad \sum_{i=1}^{r} \sum_{j \in N_i} v_{ij} x_{ij}$$

Profit from all selected pairs

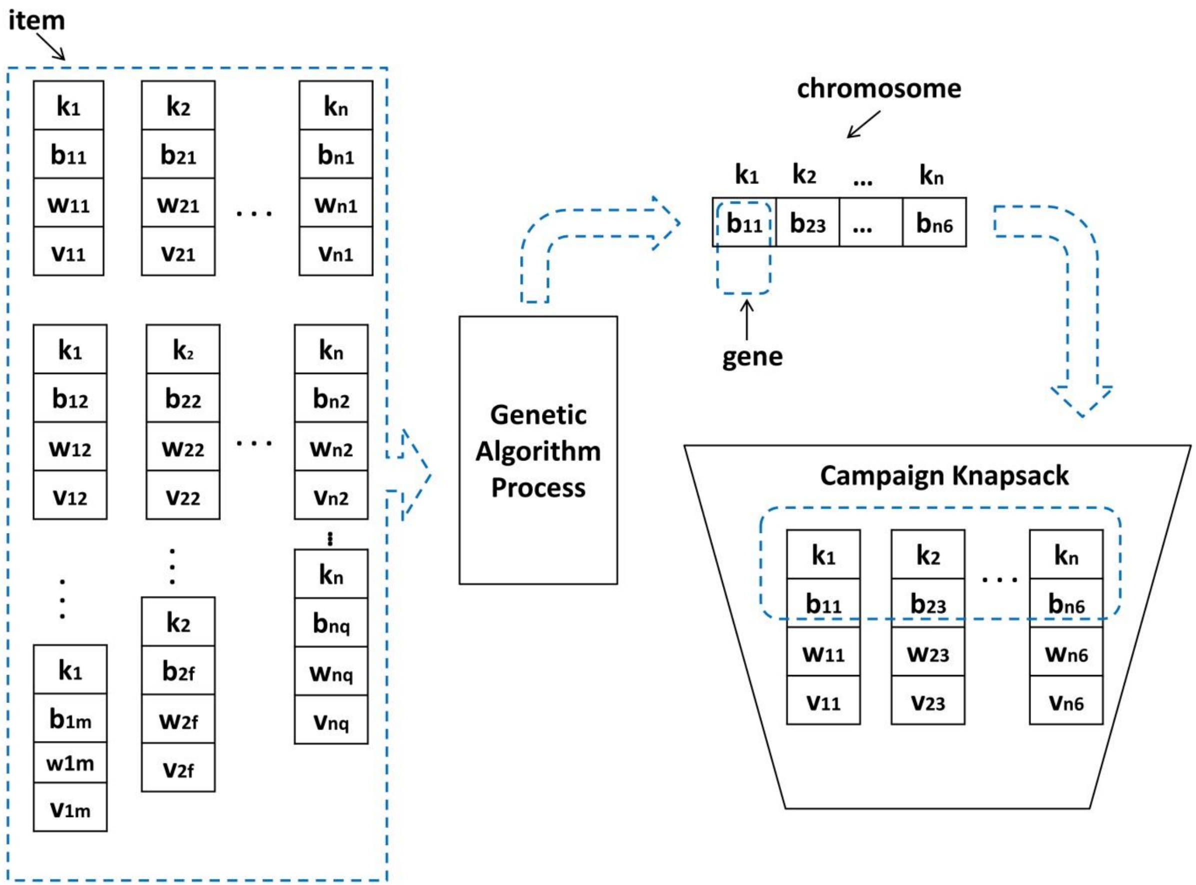
$$\text{subject to} \quad \sum_{i=1}^{r} \sum_{j \in N_i} w_{ij} x_{ij} \leq B$$

Cost of all selected pairs not exceeding budget constraint

$$\text{with} \sum_{j \in N_i} x_{ij} = 1, \text{ for all } 1 \leq i \leq r$$
$$\text{and } x_{ij} \in \{0, 1\}, \text{ for all } 1 \leq i \leq r \text{ and all } j \in N_i$$

Only one assigned bid for a specific keyword

Mapping of Campaign System to the MCKP



- Items: options of keyword-bid pairs along with their profit *v* and cost *w*
- Chromosome ≡ Set of selected items

Why Genetic Algorithm?

- Deterministic methods will always find the same approximate solution in each run – choosing persistently certain keywords
  - Adapt much slower than a method with Exploration / Exploitation
- GA: Finds an approximately optimal solution
- Stochastic approach: Selection and mutation are based on probability and randomness
- Flexibility – Adaptive Strategy
  - Discover faster changes of keywords performance

Parameters Initialization

- Keywords and Bids
  - Define a default initial bid for all keywords that are going to be tested
    - $b_{initial} = maxEstimatedFirstPageBid$
- Advertising goal
  - Optimization for traffic or monetary profit?
    1. Value = Traffic = Clicks
    2. Value = Monetary Profit = Revenue from conversions – Cost
- Begin with training periods, then optimize

Genetic Algorithm Formulation

- A possible solution is modeled as a **chromosome**

k1	k2	k3	…	kN
€0.60	€0.00	€0.45	…	€0.50

- Chromosome Fitness Function
  - Total profit expected for the bids selected in the genes
  - Evaluate the fitness *f(x)* of each chromosome *x*: $\sum v(k,b)$
  - Generated chromosome must pass the $\sum w(k,b) \leq B$ condition, otherwise randomly genes will be set to 0 until the condition is met

Optional Step: Impressions Prediction

- Alternate evaluation of the fitness function of each chromosome in the population
  - Take into consideration *predicted values* instead of *actual past ones*
- Capture **externalities** of the ad auctions and predict **current or future** behavior
  - **Global Monthly Searches (GMS)**
  - **Competition**
- Clicks, CTR, CR more dependent to inner factors (e.g., Quality Score)
- Impressions more dependent to external factors
- Multiple Linear Regression: $y' = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$
  - y: Impressions
  - x_1: Clicks, x_2: GMS, x_3: Competition

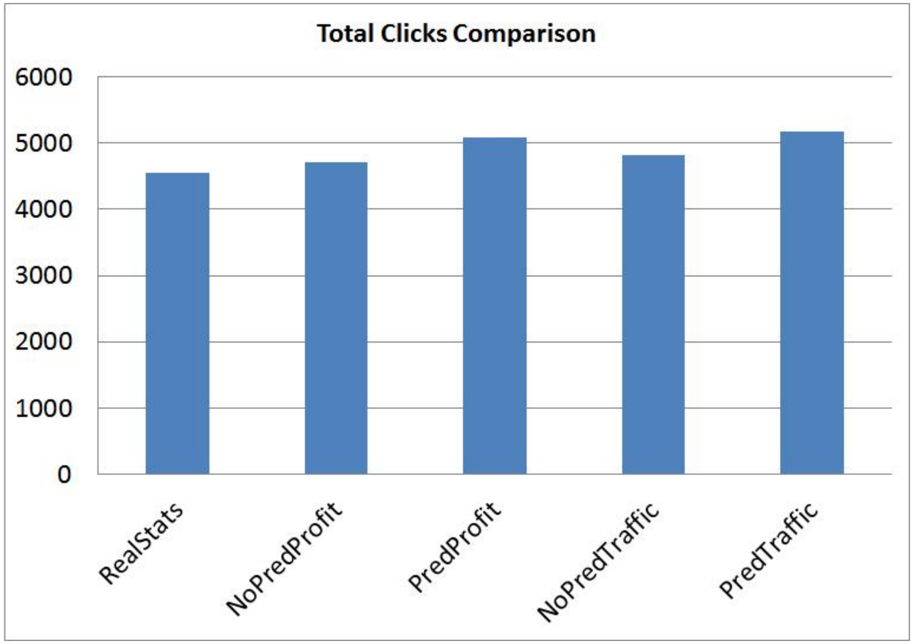
Performance Evaluation on Historical Data

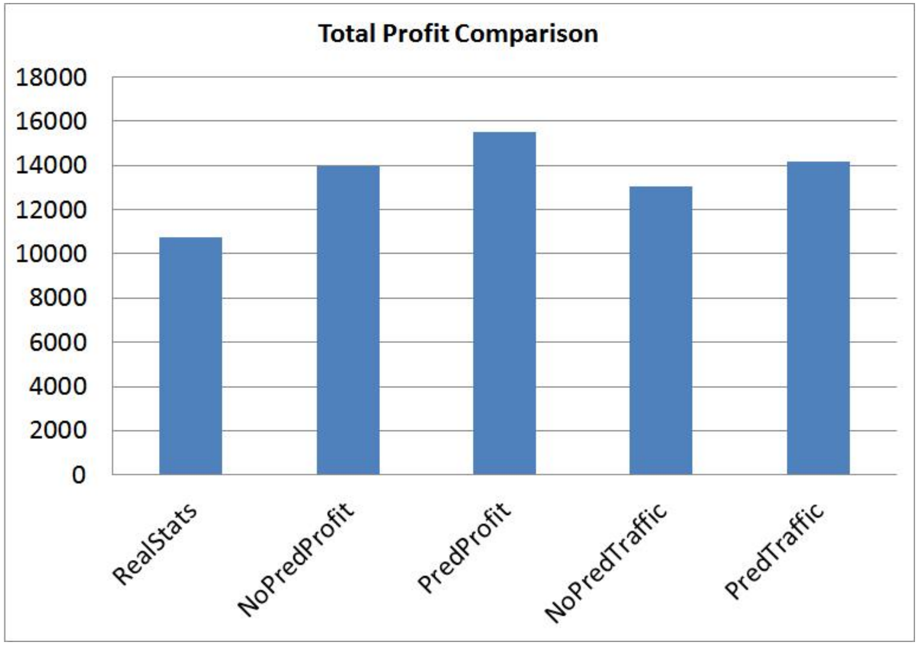
- Large scale AdWords Campaign of a website in the area of car rental – Statistics for 39 weeks

- Four basic testing scenarios:
  1. Budget Optimization for Profit with No Prediction *(NoPredProfit)*
  2. Budget Optimization for Traffic with No Prediction *(NoPredTraffic)*
  3. Budget Optimization for Profit With Prediction *(PredProfit)*
  4. Budget Optimization for Traffic With Prediction *(PredTraffic)*

- Simulation: Metrics are computed as if CTR, clicks, costs, impressions were maintained the same for each *(k,b)* choice in the future

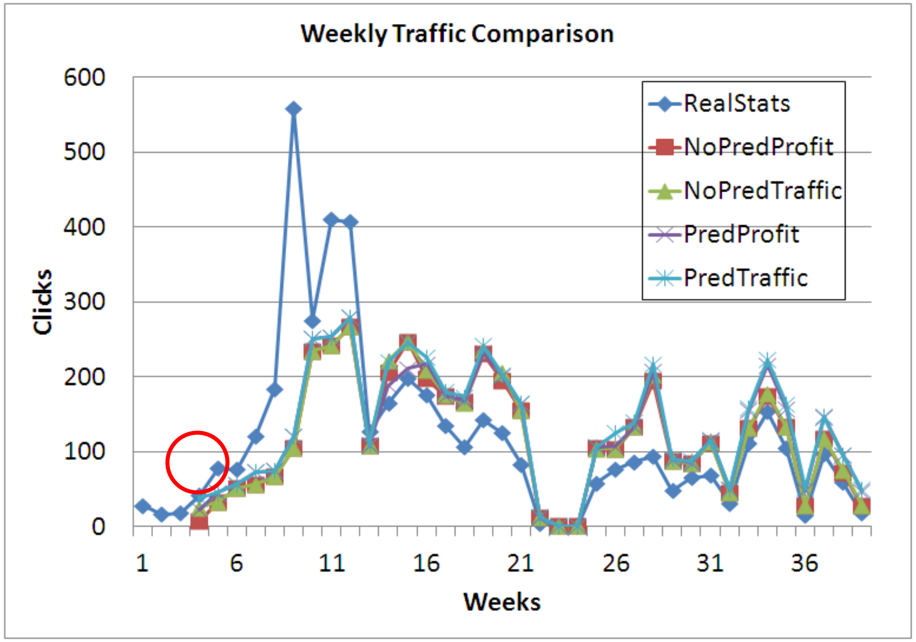
Weekly Performance Evaluation compared to RealStats

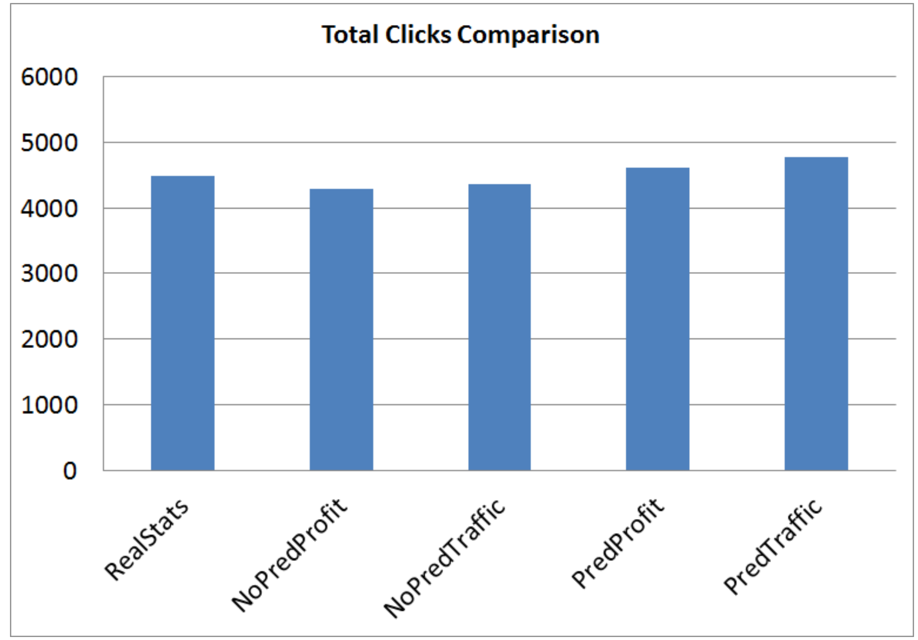




- We apply GA to evaluate the hypothesis of choosing the optimal keyword-bid combination of each week – **taking into consideration only the real used keywords and bids of the week**
- Our methods outperform the real manual bidding strategy

GA on Optimizing Next Week's Performance





- Take into consideration *(k,b)* from weeks 1 to i-1
- The advertiser until the 3rd week had been testing very few keyword options (3-4) and the GA needed more testing data to perform a valid optimization
- Our two methods **which use prediction**, outperform the real results → capture **current external factors** and conditions of the ad auction

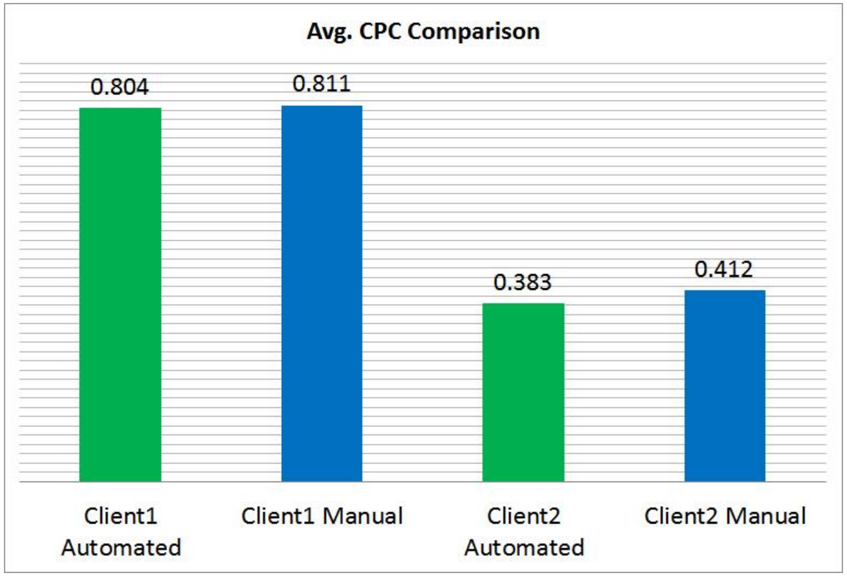
Real-time Parallel Competing Campaigns

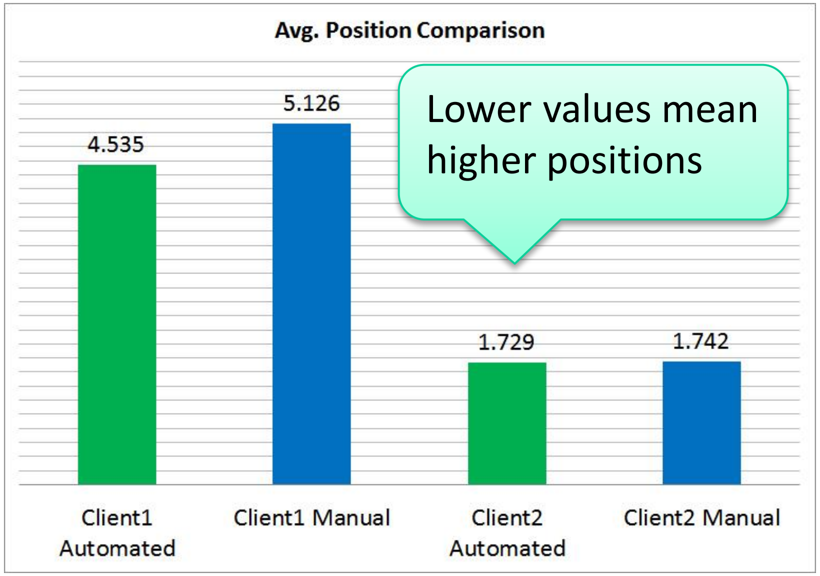
- Google AdWords campaigns for two companies
  1. Client1 is a company that offers web developing solutions (a highly competitive field for online advertising)
  2. Client2 is a company that offers aluminum railing and fencing products

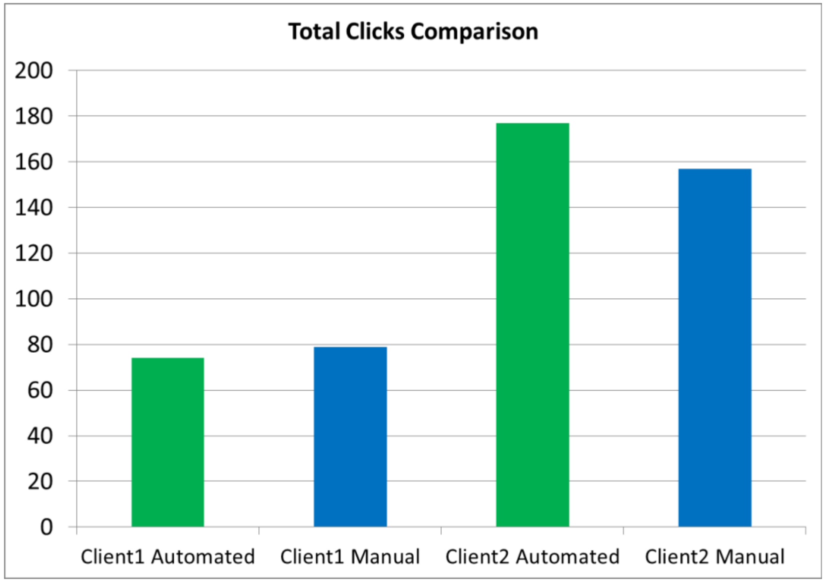
- For each company: one manual (myopic behavior) and one automated campaign
  - Advertising Goal: Optimization for Traffic
  - **Same keywords & budget** in order to test only the monitoring and optimization process
  - Active campaign period: 17 days*

* We thank **Google Greece**
for providing us coupons
In order to run the experiments

Automated Campaigns vs. Manual







Outline

- Introduction - Preliminary Online Advertising Concepts
- Keyword Generation
- Ad-Text Generation
- Campaign Optimization (Budget Optimization)
- The Adomaton Prototype
- Conclusions and Future Directions

Our Contributions

A fully implemented and functional prototype system, developed for Google AdWords platform

- *S. Thomaidou, K. Leymonis, K. Liakopoulos, M. Vazirgiannis.* ***AD-MAD: Integrated System for Automated Development and Optimization of Online Advertising Campaigns.*** *IEEE International Conference on Data Mining Workshop (ICDMW'12), Brussels, Belgium*

Experimental evaluation of the overall framework not only on a simulated environment but on real world campaigns and auction conditions as well

- *S. Thomaidou, M. Vazirgiannis, K. Liakopoulos.* ***Toward an Integrated Framework for Automated Development and Optimization of Online Advertising Campaigns****. Intelligent Data Analysis Journal. To Appear in Volume 18(6), December 2014.*

System Flow

Advertiser

Input

Main website or .CSV with landing pages

Campaign temporal length

Budget

Advertising goal

Crawler: Retrieves links from the main website

Aggregator: Imports links from the .csv file

Keyword Extraction

- Extracts bid phrases from the HTML Document

Keyword Suggestion

- Suggests extra terms using Search Engine Snippets

Ad Creative Generation

- Produces a typical sentence for an ad

System Initialization
- Campaign Settings Uploading and Parameters Organization
- Starts the Task Scheduler

Testing Period
Activates the top-relevance (untested) keywords and tests them with an initial bid for a short specified period for statistics collection

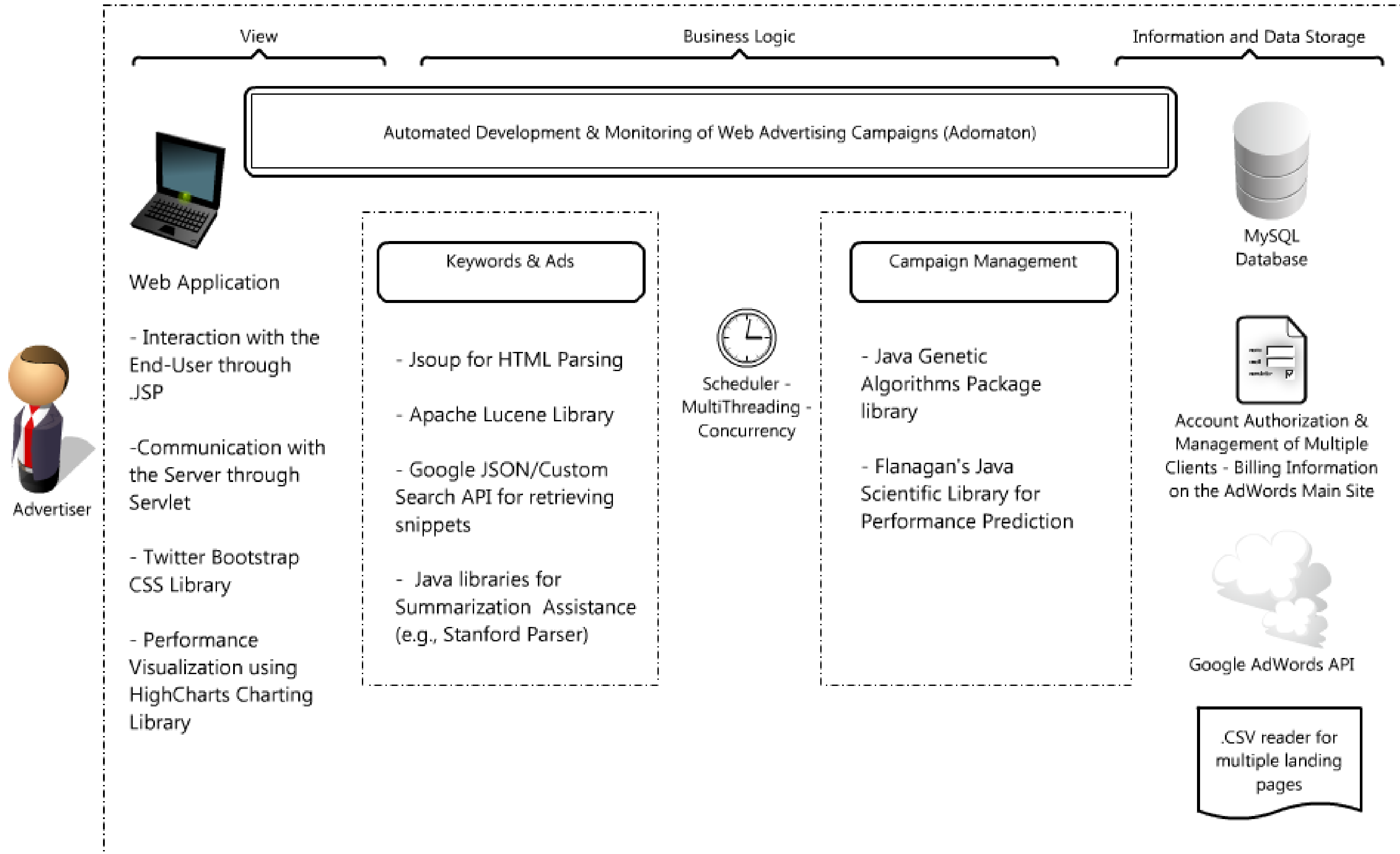
Budget Optimization
-From a set of candidate **keyword-bid** options selects with a **Genetic Algorithm** the most **profitable** combination
-Pauses the rest options
-Gives also a chance to keywords that did not receive clicks in the previous period
-Runs for a longer specified period

Optional Performance Prediction
Using **Multiple Linear Regression** and exploiting Global Monthly Searches & Competition, predicts Impressions of the next auction

Repeat the process until the Task Scheduler reaches the Campaign End Date

Focused Performance Visualization
Provides to the end user an interface which illustrates with charts the performance of key-metrics (per campaign)

System Architecture



Adomaton as a Real-time Bidding Agent

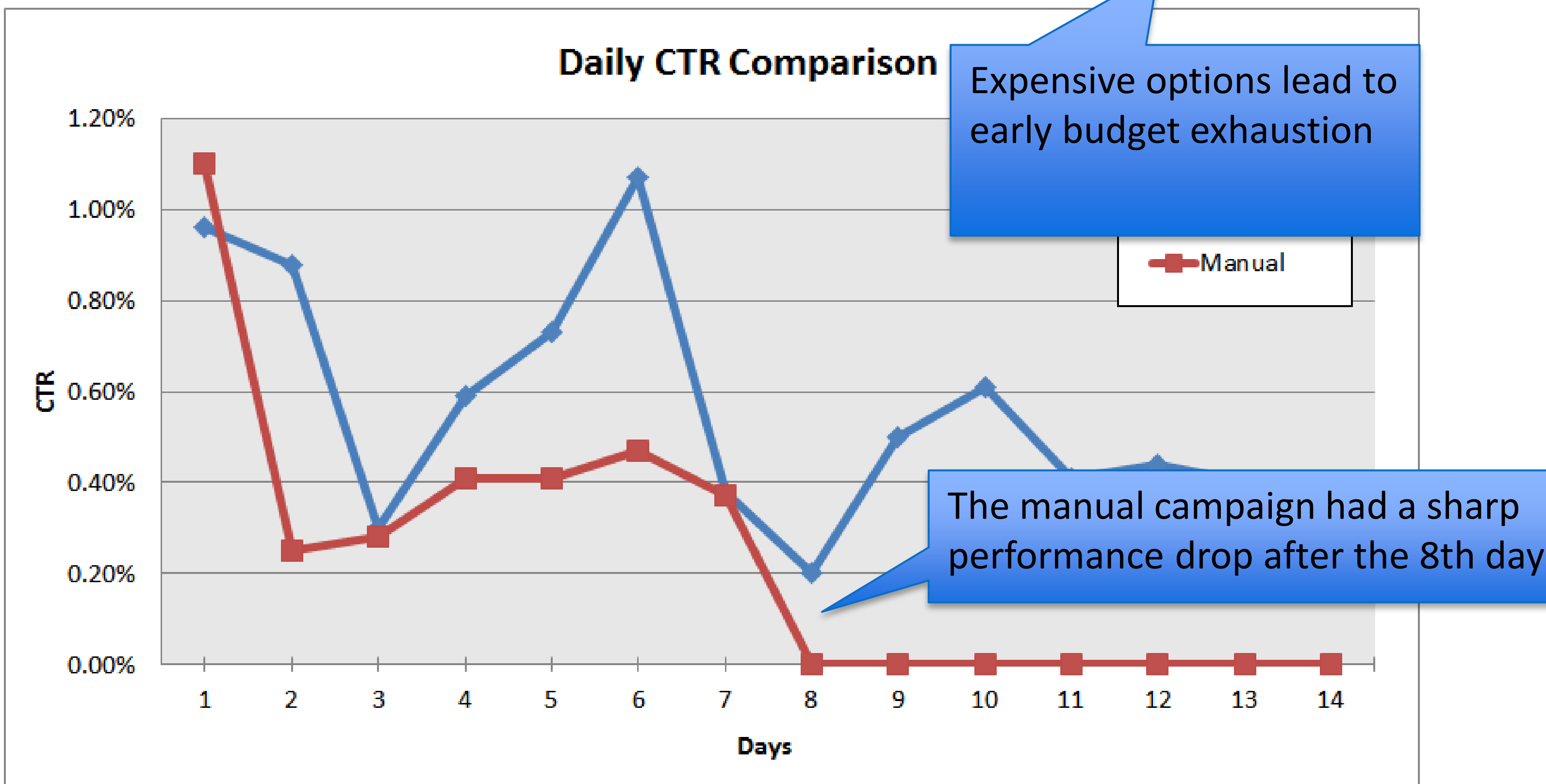
- Implementation of Main Scheduler
  - Concurrency
- Abstraction and fine granularity
- Account Authorization using OAuth 2.0 Protocol
- Prototypes
  - http://adomaton.com
  - http://prototypes-db-net.aueb.gr:8080/GrammAdsDemoDBNET

Overall Automation - Experiments

- One automated and one manual campaign of a website in the area of car rental
  - Test the **full cycle of processes** (keywords, ads, bid strategy)
  - Active campaign period: 14 days
- Manual Campaign
  - Human administrator assisted by some baseline changes from the optimization tool of AdWords
  - Myopic behavior
- Automated Campaign (Adomaton Prototype)

Results

Campaign	Clicks	Impressions	CTR	Avg. CPC
Adomaton	120	23,960	0.50%	0.92
Manual	83	21,449	0.39%	1.15



Outline

- Introduction - Preliminary Online Advertising Concepts
- Keyword Generation
- Ad-Text Generation
- Campaign Optimization (Budget Optimization)
- The Adomaton Prototype
- Conclusions and Future Directions

Contributions

- We proposed a method for recommending effective multiword advertising keywords
  - No need to capitalize on usage data such as query and web traffic logs in comparison with the state-of-the-art systems
- We proposed a novel technique which produces convincing ad-text snippets
  - It was an open problem in the SSA literature [Gabrilovich 2011]
- We proposed a profitable bidding strategy for the advertiser
  - Exploited external and current information from the ad auctions
- We developed an integrated prototype system
  - Experimentation in real-world campaign conditions and not only in a simulated environment

Future Research Directions

- Keyphrase Extraction and Creation of Ad Snippets
  - Identify automatically more product information such as price, offers, new features
  - Background and topic language model
  - Alternative Sentiment Analysis method

- Bidding Strategy
  - Prediction of clicks using boosted regression trees

System Expansion

- Web Services – Dedicated Library/API
  - Individual Modules could be used in other platforms (e.g. ad snippets for Facebook or Promoted Tweets)
  - Other external software system could be using our algorithms through API calls - e.g., a Demand-Side Platform (DSP)

Publications

1. S. Thomaidou, M. Vazirgiannis, K.Liakopoulos. **Toward an Integrated Framework for Automated Development and Optimization of Online Advertising Campaigns. Intelligent Data Analysis Journal.** To Appear in Volume 18(6), December 2014.
2. S. Thomaidou, I. Lourentzou, P. Katsivelis-Perakis, M. Vazirgiannis. **Automated Snippet Generation for Online Advertising.** ACM International Conference on Information and Knowledge Management (CIKM'13).
3. S. Thomaidou, K. Leymonis, M. Vazirgiannis. **GrammAds: Keyword and Ad Creative Generator for Online Advertising Campaigns.** Digital Enterprise Design & Management Conference (DED&M'13).
4. S. Thomaidou, K. Leymonis, K. Liakopoulos, M. Vazirgiannis. **AD-MAD: Integrated System for Automated Development and Optimization of Online Advertising Campaigns.** IEEE International Conference on Data Mining Workshop (ICDMW'12).
5. K. Liakopoulos, S. Thomaidou, M. Vazirgiannis. **The Adomaton Prototype: Automated Online Advertising Campaign Monitoring and Optimization.** Ad Auctions Workshop, ACM Conference on Electronic Commerce (AAW'12-EC'12).
6. S. Thomaidou, M. Vazirgiannis. **Multiword Keyword Recommendation System for Online Advertising.** IEEE/ACM International Conference on Advances in Social Network Analysis and Mining (ASONAM'11).

Selected References

- [Edelman et al. 2007] B. Edelman, M. Ostrovsky, and M. Schwarz. Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. American Economic Review, 97(1):242-259, March 2007.
- [Gabrilovich 2011] E. Gabrilovich: Ad Retrieval Systems in vitro and in vivo: Knowledge-Based Approaches to Computational Advertising. ECIR 2011.
- [Abhishek et al. 2007] V. Abhishek, K. Hosanagar: Keyword generation for search engine advertising using semantic similarity between terms. ICEC 2007: 89-94.
- [Ravi et al. 2010] S. Ravi, A. Z. Broder, E. Gabrilovich, V. Josifovski, Sandeep Pandey, Bo Pang: Automatic generation of bid phrases for online advertising. WSDM 2010: 341-350.
- [Joshi et al. 2006] A. Joshi, R. Motwani: Keyword Generation for Search Engine Advertising. ICDM Workshops 2006: 490-496.
- [Ganesan et al. 2012] K. Ganesan, C. Zhai, E. Viegas: Micropinion generation: an unsupervised approach to generating ultra-concise summaries of opinions. WWW 2012: 869-878.
- [Even-Dar et al. 2009] E. Even-Dar, V. S. Mirrokni, S. Muthukrishnan, Y. Mansour, U. Nadav: Bid optimization for broad match ad auctions. WWW 2009: 231-240
- [Feldman et al. 2007] J. Feldman, S. Muthukrishnan, M. Pál, C. Stein: Budget optimization in search-based advertising auctions. EC 2007: 40-49
- [Church et al. 1989] K. Ward Church and P. Hanks. Word association norms, mutual information, and lexicography. ACL 1989, pages 76–83, 1989.

Thank You

Stamatina Thomaidou’s PhD research has been co-financed by the ESF and Greek national funds through the Research Funding Program: Heracleitus II.