AD-MAD: Automated Development and Optimization of Online Advertising Campaigns

Stamatina Thomaidou*, Konstantinos Leymonis*,
Kyriakos Liakopoulos*, and Michalis Vazirgiannis*†‡
*Department of Informatics
Athens University of Economics and Business
thomaidous@aueb.gr, k.leymonis@dias.aueb.gr, mvazirg@aueb.gr, kyriakos.liakopoulos@gmail.com
†LIX, Ecole Polytechnique
‡Telecom - Paris Tech, Ecole Polytechnique

***Abstract*—Creating and monitoring a competitive and cost-effective pay-per-click advertisement campaign through the web-search channel is a resource demanding task in terms of human expertise and effort. Assisting or even automating the work of an advertising specialist will have an unrivaled commercial value. In this demonstration we present a prototype and a functional web application for semi- and fully- automated creation, monitoring, and management of cost-efficient pay-per-click campaigns with budget constraints. The prototype is experimentally evaluated on real world Google AdWords campaigns and shows a promising behavior with regards to campaign performance statistics outperforming systematically the competitive manually created and/or monitored campaigns.**

***Index Terms*—online advertising, pay-per-click advertising, automated campaign management, genetic algorithms, Google AdWords, automated keyword extraction.**

I. Motivating Automation of Online Advertising Tasks

Online advertising is gaining acceptance and market share while it has evolved into a $26 billion industry for advertisers [1] . A principal form of online advertising is the promotion of products and services through search-based advertising. Today's most popular search-based advertising platform is Google AdWords having the largest share of revenues amongst its competitors. Search remains the largest online advertising revenue format, accounting for 46.5% of 2011 advertising revenues, up from 44.8% in 2010. In 2011, Search revenues totalled $14.8 billion, up almost 27% from $11.7 billion in 2010. Web search has remained the leading format since 2006, having strong and continuous growth. The preparation of large scale online advertising campaigns for products, services, brands, or web pages can be a very complex task especially if it is designed for websites with online catalogs or catalog aggregators. The shops or listings are classified according to the products that they are selling, so each landing page contains important information and relevant description for each category or product that needs to be considered. The number of the various urls inside these domains makes the effort even more complicated regarding the manual insertion of keywords and ad-texts per landing page as well as monitoring and optimizing the performance of each campaign.

[1] http://www.iab.net/AdRevenueReport

Our proposed system aims at the automation of the mentioned tasks in order to aid the advertisers. The demonstration will present to the audience: a. keyword generation, suitable for AdWords Campaigns, from a given landing page and proposed ad creatives using text summarization, b. an automated method for budget optimization during campaign running time, based on a MCKP (multiple-choice knapsack problem) modeling and capitalizing on genetic algorithms to maximize profit or traffic, the two usual objectives for website advertising, c. a fully implemented and functional prototype system, developed for the Google AdWords platform, which currently occupies a vast share of web-search advertising volume, d. an experimental evaluation on real world data.

II. System Architecture and Components

In Figure 1, we present the overall design of the prototype and the basic components along with the associations of their inner modules.

A. System Initialization (GrammAds)

The campaigns are organized in terms of AdGroups which in turn contain keywords and ad creatives. The user initially inserts the main page for the campaign (i.e. the website url of the promoted service), the temporal legth of the campaign (i.e. days the campaign will be active), the budget amount, the goal and target of the campaign.

The three system runnable options for the campaign to be activated are the following:

1) *No Optimization*, the system just uploads automatically the generated keywords, ad-texts, and bids along with their template structure without any consideration for automated optimization.
2) *Traffic Optimization*, the advertiser considers the profit to be the amount of clicks on the ad-texts
3) *Profit Optimization*, where the profit is the actual monetary profit from offline product sales or online conversions to a specific landing page that is defined in a next step

The Crawler Module (i.e. a URL Aggregator) crawls the home page, retrieves the active links in order to discover the candidate sub-landing pages, and matches them to a

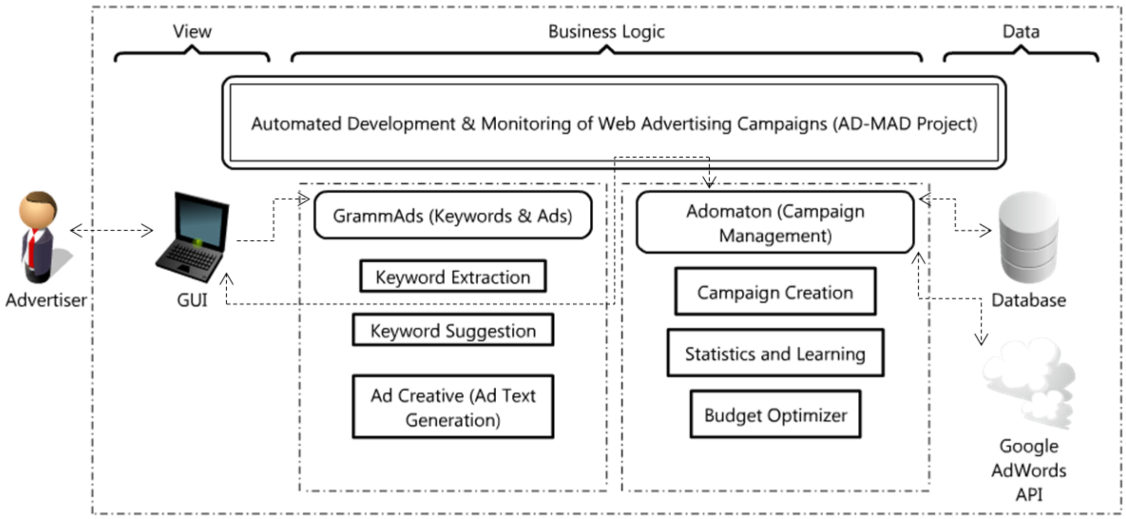

Fig. 1. AD-MAD System - Components and their Modules

corresponding AdGroup. If the user has previously selected Profit Optimization, then for each sub-landing page there is an option for the user to insert the monetary profit that he will receive from a conversion inside this url. We make the design choice that for each landing page there must be a separate AdGroup that contains its keywords and ad creatives.

1) Keyword Generation: As a third step in Figure 2, the system generates automatically keyword recommendations for each AdGroup. Next to each keyword, a normalized score of its relevance to the AdGroup is presented to the user, as well as an initial bid value. This value is derived from $min(1, estimatedFirstPageCPC)$. An initial approach of this part (focused on the keyword selection procedure) is described in [1]. In summary, we follow a corpus independent approach to rely solely on the given landing page document. We consider trigrams first, bigrams second and unigrams third from all the extracted n-grams, modifying their relevance score proportional to the number of grams. Next, we use them as seed terms for the additional suggestions For each given seed term, the keyword is submitted as a query into a search engine API. The top 30 snippet results are downloaded and loaded in Apache Lucene Library as small documents. We parse them and we construct a new vector of n-grams.

2) Ad Creative Generation: As a fourth step, the user can select or edit for each sub-landing page the automatically generated advertising text. In this subprocess the first step was to extract all the text from the HTML document of the given landing page. Then, we used summarization to keep the most important meaning for the description of our advertising page, resulting in a text summarization of one sentence -which is considered as the most important one retrieved from the landing page- in order to insert it into the description lines of the ad creative. In the end of the second description line we add a call-to-action phrase.

B. Initial Testing Periods gathering Statistical Information

A SQL database, must store all the statistical information collected for the keywords to be able to track their performance. All information about keywords are necessary for the budget optimization module function. We also store temporal information, in the sense of distinguishing time periods. Also, there is the need for data access objects (DAOs) that retrieve and insert data to the database in a simpler manner. Both the Genetic Algorithm and the Prediction modules make extensive use of the statistical information of the keywords.

We need to define a default initial bid for all keywords that are going to be tested, so given a specific variable information from Google AdWords at the Keyword Generation Step, we set $b_{initial} \leftarrow maxEstimatedFirstPageBid$. Next, we define time intervals for task periods (e.g. 2 days in our experiments). The general form of training periods to test the initial keyword performance without optimizing is the following:

- For a subset of the most relevant n keywords of each adgroup for testing (bid $b_{initial}$) and collect statistics.
- Let a new subset of the next most relevant n keywords of each adgroup for testing (bid again $b_{initial}$). Make a random change in $m \ll n$ keyword bids from the previous subset: ($b_{new} = b_{previous} \pm b_{previous} \times 50\%$). Collect statistics.
- Perform Optimization

The genetic algorithm step is the implementation of the optimization process. Finally, in each testing phase after optimization, we follow the same process of the first testing periods but we also pause previous keywords that are not selected by the optimization module.

C. Optimization and Monitoring (Adomaton)

The statistics module result in a list of all possible keywords. Each keyword of the list must have all the possible bids for which we have statistics. For each keyword-bid combination

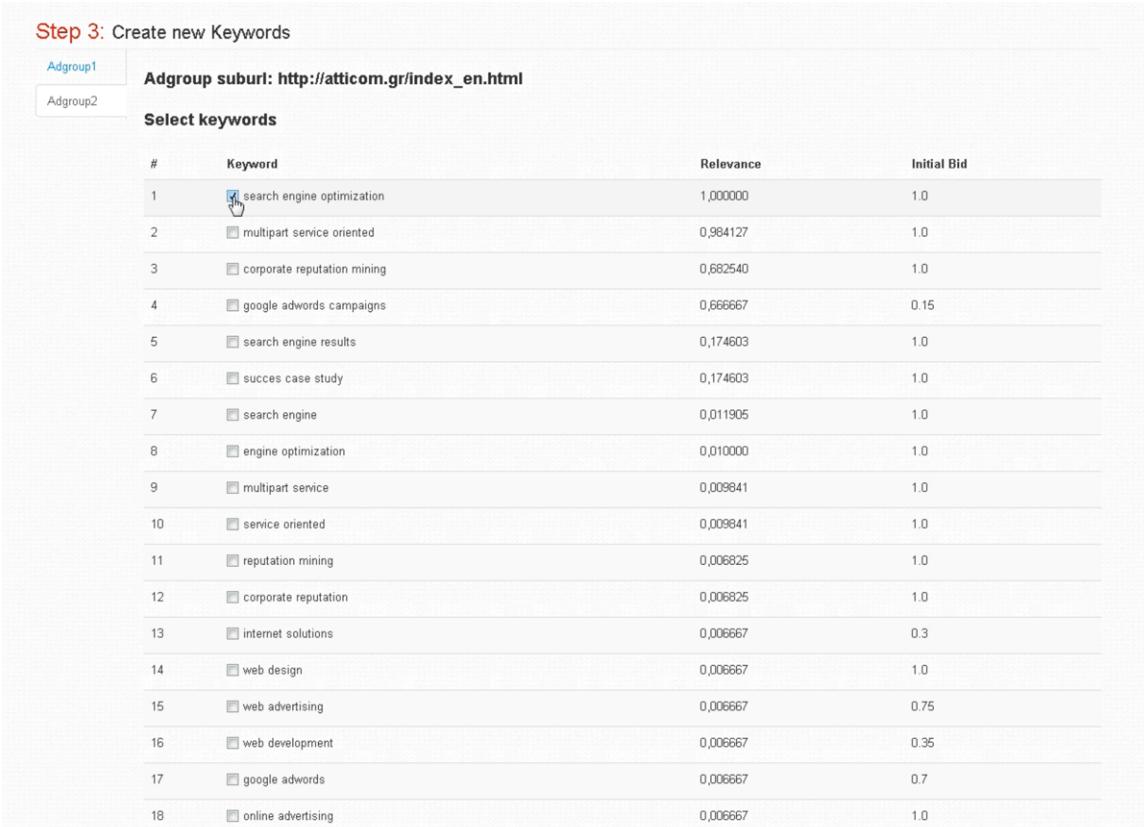

Fig. 2. Presentation of Keywords-Relevance-Proposed Bids

(k, b), an evaluation of the cost and profit must be computed based on the keyword performance resutls kept in the database. Optionally, instead of directly using the number of keyword impressions to compute cost and profit, prediction can used. The problem is then modeled into chromosomes and the fittest chromosome is finally selected by a *Genetic Algorithm* after several generations of breeding using the methods of crossover, mutation, and elitism. A final list of keyword-bid pairs (k, b) is produced from the fittest chromosome. These keyword-bid pairs will form the new *bidding strategy*, which must be set in the AdWords account of the advertiser using the AdWords API. An initial approach focused on the budget optimization process along with evaluation and several experiments is described in [2].

The optional *Prediction Module* aims at predicting impressions using past statistics of keywords and targeted information taken from AdWords, such as the *Global Monthly Searches* and the level of *Competition* that exists for a given keyword. Before budget optimization, optionally, we can use this module in order to refine the statistics used by applying impressions prediction. A *multiple regression model* finds the relationship between the impressions and other variables, such as clicks, global monthly searches, and competition.

III. Real-time Parallel Competing Campaigns

In order to test the performance of our system in real-world data, we create Google AdWords campaigns for two industrial web sites; Client1 offers web design solutions (a highly competitive field for online advertising) while Client2 web site offers aluminum railing and fencing products. For each client we create a manual and an automated campaign. Each automated campaign is created semi-automatically by our system (the only intervention is the parameter input of daily budget, account credentials, period of active campaign, and keywords). We set our automated campaigns for traffic maximization as the advertising goal. We use for each manual and automated campaign the same keywords and budget in order to have comparable results with regards to monitoring and optimization. In this experiment, we do not use impressions prediction, only the real values case scenario (due to limited budget for further experiments at that time). In Figures 3 and 4, we present the final results after a period of 17 days. In the case of Client2, the automated campaign achieved higher performance in total traffic than the manual one. In the case of Client1, the automated achieved a slightly lower performance than the manual one. In both cases, the automated campaigns achieved *better placement in the advertising slots* than the manual ones, as well as *lower prices for average cost-per-click*.

IV. ADMAD demo URL

A demonstration of the system can be found in the following screencast: http://www.adomaton.com/. Interested audience can also experiment with the GrammAds component

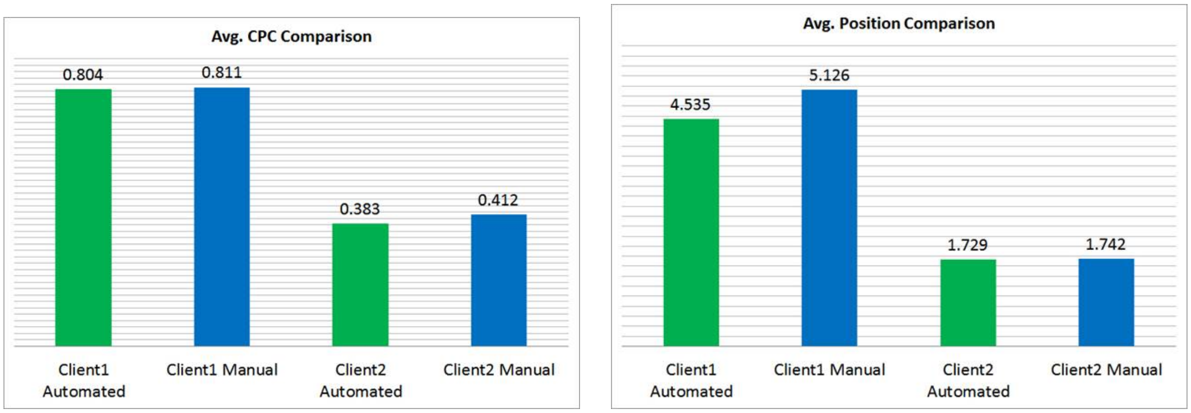

Fig. 3. Automated compared to Manual Campaigns

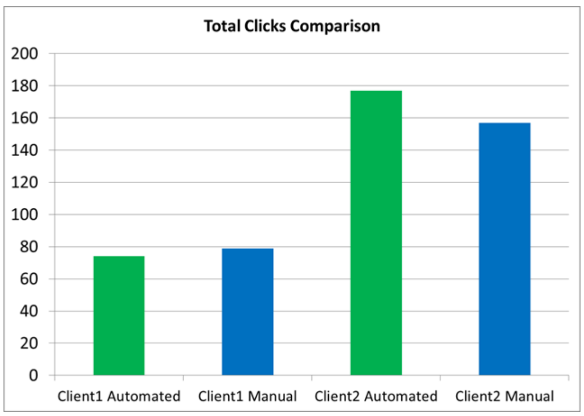

Fig. 4. Total Clicks Comparison

in http://www.grammads.com/ that proposes keywords and an experimental form of ad-texts.

V. Conclusion and Further Discussion

Our contributions regarding the improvement of the advertising campaign development process consist in:

- Automating the task of finding the appropriate keywords for an advertising campaign
- Recommending multiword terms (n-grams) with high specificity without the need to capitalize on usage data such as query and web traffic logs
- Generating snippets of ad texts
- Proposing a method of overall campaign optimization
- A fully developed system with convincing experimentation on real world data from various thematic areas

The exploitation of a genetic algorithm and also the impressions prediction method for this type of problem form innovative solutions with respect to existing literature. Regarding the automated ad creative generation process, to the best of our knowledge, this issue remains still an open problem in natural language processing and information retrieval areas as mentioned in [3].

Currently, we are working towards alternate bidding strategies and prediction of clicks using regression trees with some preliminary but promising results. We estimate for our next steps to complete the migration of Campaign Behaviour Prediction Module into the integrated system. A further extension on our system can be the improvement of the ad creative generation component. The creation of specialized ad text will be based on previous work and research studies on paraphrasing methods and text summarization. In combination with category specific templates which will be filled with the product characteristics, such as name, price, location, etc., the system will generate ad text for the advertisements of the campaign. The above features will be extracted from customers web page primarily. Regarding the overall system performance evaluation, we aim to conduct larger and combined experiments, testing extensively and concurrently the application for multiple users and campaigns for long periods of time.

Acknowledgment

The research of S. Thomaidou is co-financed by the European Union (ESF) and Greek national funds via Program Education and Lifelong Learning of the NSRF - Program: Heracleitus II. Prof. M. Vazirgiannis is partially supported by the DIGITEO Chair grant LEVETONE in France.

References

[1] S. Thomaidou and M. Vazirgiannis. Multiword Keyword Recommendation System for Online Advertising. *In Proceedings of the 2011 International Conference on Advances in Social Network Analysis and Mining (ASONAM '11), Kaohsiung, Taiwan. IEEE Computer Society*

[2] K. Liakopoulos, S. Thomaidou, M. Vazirgiannis. The Adomaton Prototype: Automated Online Advertising Campaign Monitoring and Optimization. *In Eighth Ad Auctions Workshop, 13th ACM Conference on Electronic Commerce, June 2012, Valencia, Spain*

[3] E. Gabrilovich. Ad Retrieval Systems in vitro and in vivo: Knowledge-Based Approaches to Computational Advertising. *In ECIR 2011*