GrammAds: Keyword and Ad Creative Generator for Online Advertising Campaigns

Stamatina Thomaidou, Konstantinos Leymonis, Michalis Vazirgiannis

Abstract Online advertising is a fast developing industry - in 2011 its revenues reached $31 billion. Paid search marketing is extremely competitive while online advertising campaign creation and development are very demanding in terms of time and expert human resources. Assisting or even automating the work of an advertising specialist has emerged as a requirement for companies and research institutes over the last few years, mainly because of the commercial value of this endeavour. In this context, we developed *GrammAds*, an automated keyword and ad creative generator. This system generates multiword keywords (n-grams) and automated ad creative recommendations, while it organizes properly the campaigns which are finally uploaded to the auctioneer platform and start running. In this paper, we analyze the proposed methodology and we also present the main functionality of the *GrammAds application* along with an experimental evaluation.

1 Introduction

Online advertising is gaining acceptance and market share while it has evolved into a $31 billion industry for advertisers [1]. One form of online advertising is the promotion of products and services through search-based advertising. The three most prevalent options in the search-based advertising market are Google AdWords, Ya-

Stamatina Thomaidou
Athens University of Economics and Businees, e-mail: thomaidous@aueb.gr

Konstantinos Leymonis
Athens University of Economics and Business e-mail: k.leymonis@dias.aueb.gr

Michalis Vazirginnis
LIX, Ecole Polytechnique & Athens University of Economics and Business
e-mail: mvazirg@lix.polytechnique.fr

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

hoo Search Marketing, and Microsoft AdCenter (the two latter have merged) [2]. 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 revenues, up from 44.8% in 2010. In 2011, Search revenues totaled $14.8 billion, up almost 27% from $11.7 billion in 2010. Search has remained the leading format since 2006, having strong sequential growth. [3].

Baseline Campaign Creation Process: To create an advertising campaign usually the advertiser must have one or more products on a website that wants to be exposed to the public. For each product there must be a landing page, which is the webpage a user will be redirected to, after clicking the advertisement of the product. The landing page is usually the place where the user can see information about the product, its technical characteristics, its price, and has the option to buy it. After finding what the advertising wants to sell and preparing the landing pages, it is crucial to select on which keywords (words or phrases) each product will be advertised. The keywords used for each product must be relevant to it, otherwise the campaign will not be profitable. A good practice is to choose the most specific key-phrases possible, which usually consist of 1-3 words.

After finding the keywords, the advertisement texts (or ad creatives) must be correctly written. They must be short and precise, understandable with convincing calls to the user to take action. Ad creatives consist of a short headline, two limited lines of description and a display url, which does not have to be the same as the real landing page url. Keywords and ads belong to AdGroups. The ads of an AdGroup are shown for keywords belonging to the same AdGroup. Thus, it is important not to mix ad-texts and keywords of irrelevant products in the same AdGroup. A budget must also be set on every campaign which will be consumed. An advertiser must also decide how much the maximum cost-per-click of each keyword will be. This is the bid that the advertiser is putting for a keyword and approximately an upper limit of how much each click for this keyword may cost.

Motivation for an Automated Application: As it is implied, effective *keyword selection* is one of the most important success factors for online advertising. Companies would like to advertise on the most effective keywords to attract only prospective customers and not uninterested browsing users, while they are also in need of well-written ad creatives to attract more visitors and generate thus higher revenues. In addition, the preparation of large scale online advertising campaigns for products, services, brands, or web pages can be a very complex task if it is designed for websites with *online catalogs* or catalog aggregators (e.g. http://www.fnac.fr/). 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

[2] http://www.searchalliance.com/publishers

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

insertion of keywords and ad-texts per landing page. Our proposed system aims at the automation of the total procedure in order to aid the advertisers.

2 Related Work

The manual selection of even a small set of keywords for advertising purposes is quite laborious, a fact which leads to the recent appearance of commercial tools that create keyword sets directly from a landing page. There exist different techniques for keyword generation. Search engines use query log based mining tools to generate keyword suggestions. In this way, they focus on discovering co-occurrence relationships between terms and suggest similar keywords. They start from an initial key phrase and they are based on past queries that contain these search terms. Google AdWords Keyword Tool [4] exploits this ability and presents frequent queries for the seed set of terms

Other commercial tools [5] determine an advertiser's top competitors and then actively search for the keywords they are targeting. After a period of time, lists of targeted keywords that are competitive for pay per click advertising are automatically generated. These two approaches may result to a recommendation set of keywords which are likely to be general and thus more expensive. Considering this, the challenge of generating keywords is to select both semantically similar and well-focused keywords.

TermsNet and Wordy [8, 1] exploit the power of search engines to generate a huge portfolio of terms and to establish the relevance between them. After selecting the most salient terms of the advertiser's web page they query search engines with each initial seed term. With their methods they find other semantically similar terms. Wordy system proposed single word terms (unigrams) for each seed keyword. S. Ravi et al. [12] propose a generative model within a machine translation framework so the system translates any given landing page into relevant bid phrases. They first construct a parallel corpus from a given set of bid phrases b, aligned to landing page keywords l, and then learn the translation model to estimate $Pr(l|b)$ for unseen (b,l) pairs. This approach performs very efficiently but depends on the chosen domain and data that the human decision factor may affect.

In general, *corpus or domain dependent* systems require a large stack of documents and predetermined keywords to build a prediction model [11], while on the other hand our developed system works with a *corpus independent* approach that directly sifts keywords from a single document without any previous or background information.

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

[4] http://www.adwords.google.com/keywordtool

[5] http://www.adgooroo.com/, http://www.wordstream.com/

Information Retrieval areas as mentioned in [6]. Thus, the corresponding module of our system is an innovative contribution in this regard.

3 Keyword Generation

This component aims at proposing valid and representative keywords for a landing page capitalizing on keyword extraction methods, on the co-occurrence of terms, and on keyword suggestions extracted from relevant search result snippets. [6] A more detailed presentation of this component is described in [14]. In the following two paragraphs we will describe briefly their main functionality.

Keyword Extraction: In this process, we follow the corpus independent approach to rely solely on the given landing page document. As a preprocessing step, the HTML content of each landing page is parsed, stopwords are removed and the text content is tokenized. Next, for each word (unigram) in the tokenized output, we compute a special tag weight. From the most retrieved unigrams we pull together possible combinations of two-word phrases (bigrams) inside the given landing page. Following the same co-occurence discovery process we extract three-word terms (trigrams) as well. By gathering all terms, we construct the final extracted keywords vector. We boost trigrams first, bigrams second and unigrams third, modifying their relevance score proportional to the number of grams.

Keyword Suggestion: From the previous step of keyword extraction we have already extracted the initial keywords. These will be the seed keywords for the additional suggestions For each given seed keyword, 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 [7] as small documents. We parse them and we construct a new vector of grams. Based on the Lucene scoring method we find the unigrams and bigrams that have the largest number of occurrences inside the document and thus are kept as the most relevant for the specific seed query. Each of these terms is representing a new query. Again, we sort in descent order the new queries based on this score and we create a vector of suggested keywords and their scores for each of the seed terms.

4 Ad Creative Generation

Here, we propose an automated process for the *ad creative generation*. Summaries of Web sites help Web users get an idea of the site contents without having to spend time browsing the sites. The technology of automatic text summarization is maturing and may provide a solution to the information overload problem. Automatic text

[6] A demonstration of the process can be found in the developed web application at www.grammads.com

[7] http://lucene.apache.org/java/docs/index.html

summarization produces a concise summary of source documents. The summary can either be a *generic summary* (this type of summary we using in the next process), which shows the main topics and key contents covered in the source text, or a *query-relevant* summary (which is a further challenge of our system as we could use specific keywords for filtering the resulted summaries), which locates the contents pertinent to user's seeking goals [15]. 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. For this purpose the input was the text from the page to the Classifier4J [8]. Using a Bayesian classifier implementation, we can get 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 such as: "Buy now!", "Purchase now!", "Order now!", "Browse now!", "Be informed" according to each advertising goal. For example we use "Be informed" for advertising a web page or a brand name, "Buy now!" for advertising a product, "Purchase now!" for advertising a service. The purpose of adding these phrases into the text ad is to optimize the ad descriptions because a call-to-action encourages users to click on the ad and ensures they understand exactly what the advertiser expects them to do when they reach the landing page.

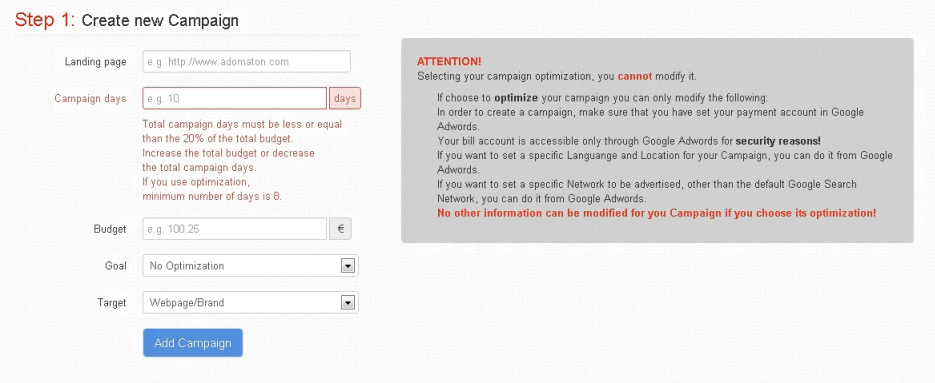

Fig. 1 Initialization Settings

5 Campaign Organization and Use Cases

The main objective of the system developed is to improve the user experience of setting up and optimizing small or large scale Google AdWords Campaigns, no matter

[8] http://classifier4j.sourceforge.net/

how experienced the user is, in a very simple, straight-forward way, minimizing the needed input information and the total steps of the process. In that way, GrammAds application significantly reduces time, resources, and final effort of the advertisers. For a mid-size one product campaign (i.e. 3 Adgroups - 30 keywords) the labour cost for a manual campaign is $\sim$ 30 person hours per month, whereas using the proposed system the whole process drops to a few minutes for inserting the desired landing pages and the basic goals of the campaign.

In the following paragraphs we will present a basic example of the application use. In each case, user has to have been logged in the system, with the AdWords credentials. Adomaton Servlet starts a new session for this user, assigning all the information needed between requests and responses, to an object, which then is being saved in this specific session as attribute. User selects creating a new AdWords campaign and he is directed to the page that is presented in Figure 1. User fills the information needed for the campaign that the application is going to create, inserts the main URL of the product, service, or brand-name that he wants to promote, the period days of the whole campaign, and the budget for the total campaign days. In addition, he selects one of the three system runnable options for the campaign that is going to be activated:

1. *No Optimization*, where the system just uploads automatically the generated keywords, ad-texts, and bids along with their organized structure without continuing to be responsible for an automated optimization strategy
2. *Traffic Optimization*, where the advertiser considers the profit to be the amount of clicks at 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

Then, user selects the advertising target that can be a.Website/Brand-name, b.Product, c. Service. This option is useful for the Ad Creative Module in order to generate the proper action phrase. The default option from our system regarding the Google Network where the ad-texts are going to be impressed is only the group of *Search Network*, opting-out the same time from the Display Network group. We took this decision in our experiments and strategies because choosing to appear also in the Display Network was leading to a large amount of impressions and very few clicks. As a result the values of CTR (Clickthrough rate) were very low ($< 0.5\%$) causing in this way low Quality Scores and increased recommended bids for good ad slots.

After the proper settings are inserted, Adomaton Servlet as a second step reads each input information and assigns them to the session object. Depending on the main landing page that has been set by the user, the *Crawler Module* visits the specified webpage and using a web-scraping technique, obtains webpage source information. In this process, the Crawler extracts possible existing sub-landing pages, after validating their availability. Each retrieved sub-landing page is corresponding semantically to an AdGroup in our proposed Campaign Organization. For the case of Profit Optimization, next to each sub-landing page the user can insert a specific monetary profit that he is going to gain from a conversion in this page. This is useful

for the Profit Maximization Strategy (a detailed algorithm is presented in [10]). The user can select all or a portion of these retrieved landing pages.

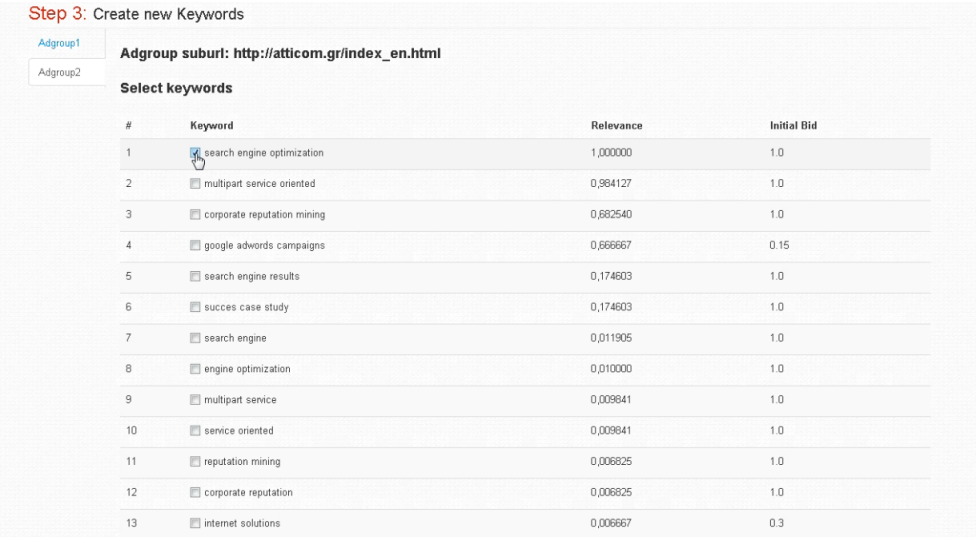

Fig. 2 Automatically generated keywords and recommended bids

As a third step, the user is directed in a page where he must select for each of the previously selected pages the automatically generated keywords. Next to each keyword it is presented to the user a normalized score of its relevance to the AdGroup, as well as an initial bid value. This value is derived from $min(1, estimatedFirstPageCPC)$. This step is presented in Figure 2. The estimated first page bid amount approximates the cost-per-click (CPC) bid needed for the ad to reach the first page of Google search results when a search query exactly matches the keyword. This estimate is based on the Quality Score and current advertiser competition for that keyword. We retrieve this value using the AdWords API [9]. We decided to place an upper bound of 1 due to the fact that a commonly used strategy by the advertisers is to evaluate as a default case the utility of the click at a url to be equal to one monetary unit (e.g. 1 euro or dollar) [4, 2]. [10]

Furthermore, the user at the fourth step in Figure 3 can select for each sub-landing page the automatically generated advertising text as well as edit each part of it to his needs, before its final insertion in the campaign.

Finally, in Figure 4 the user can see through the AdWords interface the uploaded settings of his constructed campaign.

[9] https://developers.google.com/adwords/api/

[10] http://support.google.com/adwords/bin/answer.py?hl=en&answer=2471184&from=50081&rd=1

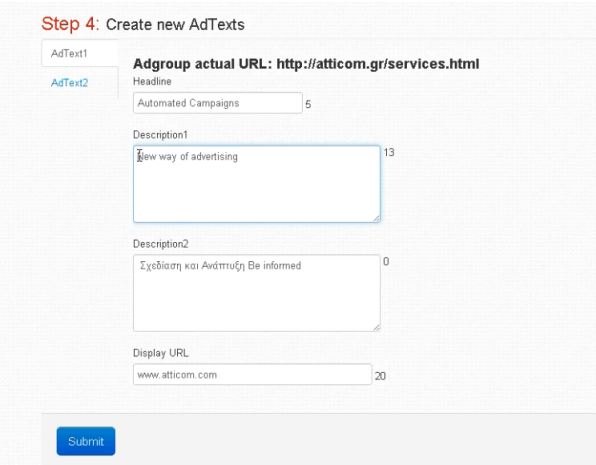

Fig. 3 Automated Ad Creatives

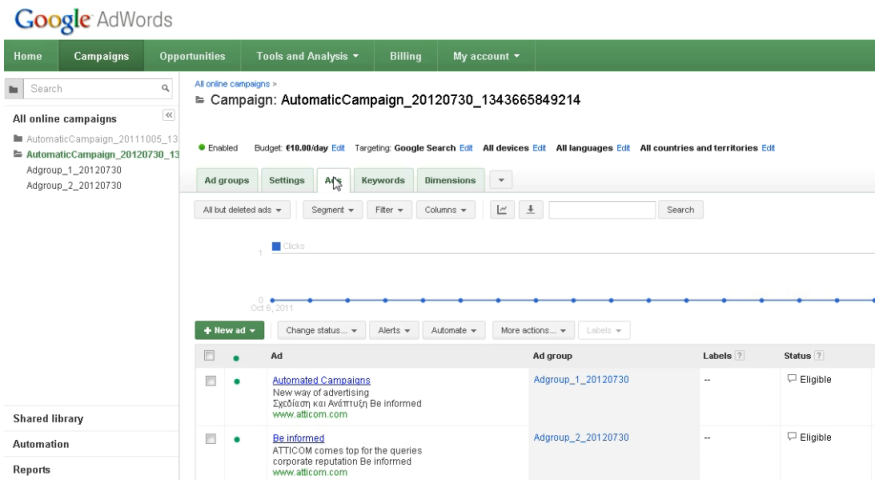

Fig. 4 AdWords Uploaded Settings

6 System Architecture and Communication

The proposed system structure resulted as a need for an appropriate user interface interaction of the *Adomaton* subsystem, which optimizes Google AdWords Campaigns through a novel bidding strategy. In this logic, it is intended to allow users access the system through a web interface, available from their browsers, just by logging in a webpage, using their registered account's AdWords credentials. Each registered Google AdWords user, can use the system as a client. Our web application is equipped with its own server and database, in order to keep the needed information such as statistics and perform the initial upload of the campaign in the

corresponding user account on AdWords API, optimize and finally monitor the campaign's performance.

For the development of the interaction system, Servlet is used, a Java-based server-side web technology, which extends the capabilities of a server that host applications access via a request-response programming model. This technology is tied with the HTTP protocol in order to comply with the web browser access to it. The main reason that Servlet has been used is because of the optimization algorithms that have been previously developed and tested with the Java programming language. In order to display the content of the information Model-View-Controller (MVC) is used as a pattern for the information management. More specifically the Java Server Pages (JSP) technology were included in order to dynamically generate web pages based on HTML markup notation language. To deploy and run Java Server Pages , it was required Apache Tomcat, a compatible web server with a Servlet container. AdWords API has been used in order to ensure the connection of the system with the AdWords system, and thus use the AdWords modules to synchronize each campaign's options and settings with our system. The development methodology that has been used is the incremental development [9], a combined linear-iterative framework. In that way, it was performed a series of waterfalls consisting of requirements extraction, design, implementation, testing and maintaining of the system. An important aspect that we took under consideration are the constant AdWords API changes and migrations. Thus, in need of a more proper code organization and maintenance, we constructed a package of *AdWords API Wrapper*, to keep up with all the corresponding library major edits.

In Figure 5, we present the *GrammAds* architecture with the associations of the basic modules. This system was developed in the context of an *overall automated solution* for creating, monitoring, and optimizing a Google AdWords campaign. A detailed description of the *Budget Optimization* process and main strategy can be found in [10], as well as a beta demo version in [13].

7 Experimental Evaluation

In our initial experiments described in [14], we evaluated at a first glance the recommended keywords using human ranking following a blind testing protocol. In the next steps we evaluated also the automatically generated ad-texts. Eleven researchers and postgraduate students provided feedback in the scale of Grade.1:Bad up to Grade.5:Very Good. In Figure 6 we present the evaluation results. In the context of various experiments where we used the exact same bidding strategy for two identical campaigns of a company that offers web developing solutions (a highly competitive field for online advertising), we discovered the following: The keywords that were generated from GrammAds *achieved higher Clickthrough rate (CTR) values* than the manual inserted ones as shown in Figure 6.

As a future empirical evaluation regarding the return-on-investment (ROI) of the online advertising campaigns we plan to make comparisons of the website traf-

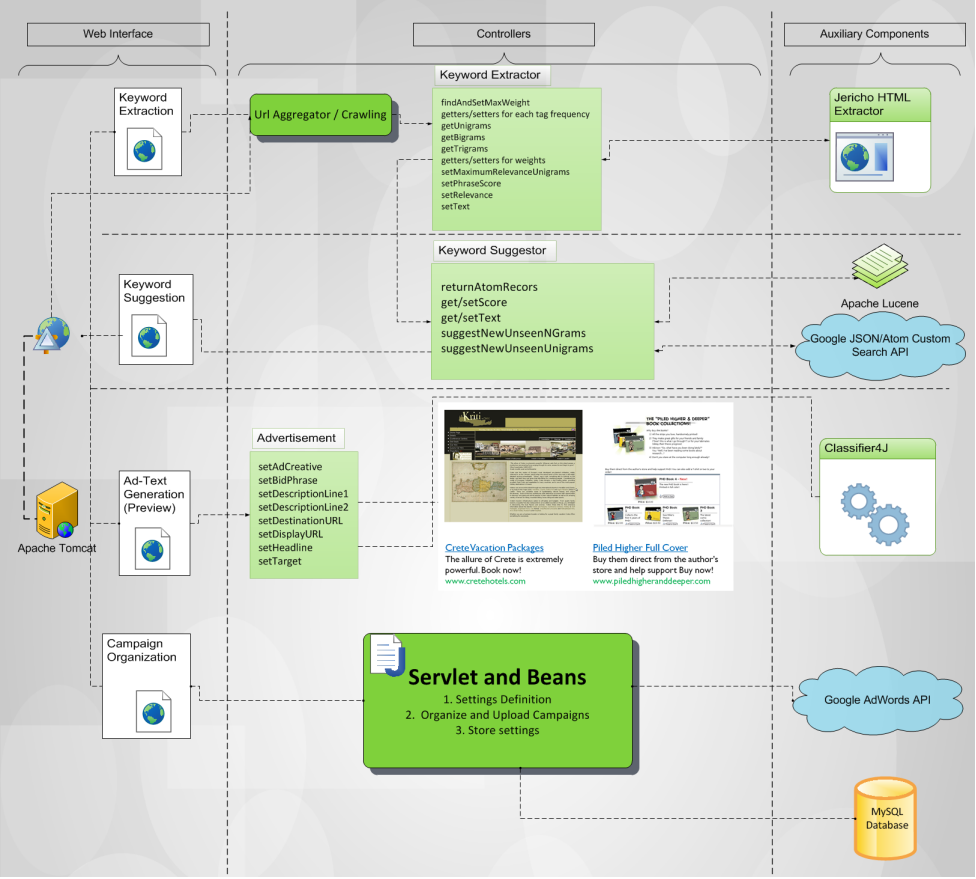

Fig. 5 System Architecture and Modules Communication

fic from organic-search sources compared to the paid-search sources along with their matched search query [11]. In order to make an evaluation more focused on the quality of the recommended keywords and ad creatives by the GrammAds application rather than focusing on more complex and sophisticated bidding strategies, we will set each day as a bid value the recommended initial bid value of our system $(min(1, estimatedFirstPageCPC))$.

8 Conclusions and Future Work

In this paper we proposed a system that, given a landing page in the context of online advertising for products and services promotion, automatically extracts and

[11] http://www.google.com/analytics/

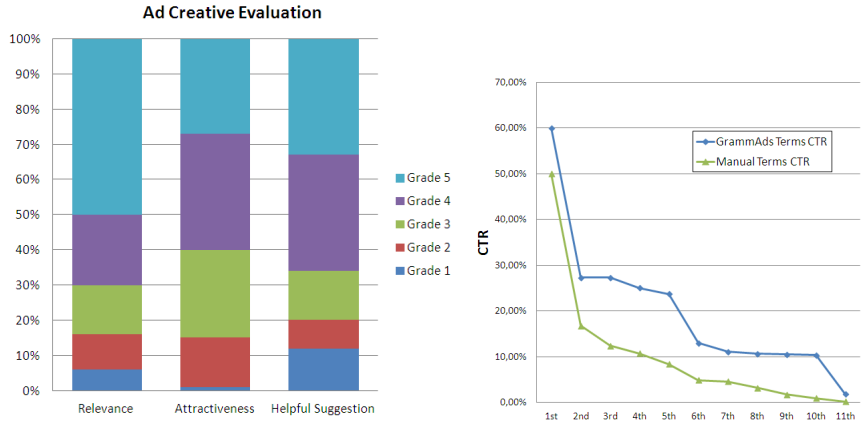

Fig. 6 Ad Creatives Feedback & Keyword CTR Comparison for top-11 generated terms

suggests keywords for web advertising campaigns as well as automatically generates advertisement texts.

In this way, our contributions regarding the improvement of the advertising campaign creation process consist in:

- Automating the task of finding the appropriate keywords
- Recommending multiword terms (n-grams) with high specificity without the need to capitalize on usage data such as query and web traffic logs
- Generating fast snippets of ad texts
- An automated and fast way of uploading campaign and AdGroup settings (keywords and ad-texts per AdGroup) into the AdWords service
- A developed web application with an initial experimentation on real world campaigns

Using the search result snippets for the process of keyword suggestion has helped a lot to retrieve faster the proper information rather than crawling actual documents. It was also a helpful mean to keep the trends and thus retrieving trending topics at a specific time. Also, searching result snippets from queries on twitter search and tags can be helpful due to the compact nature of twitter messages. They can help in filtering out irrelevant or general information, while mining market trends. A further extension on our system can be the expansion of the ad creative generation component. The creation of specialized ad text will be based on previous work and research studies on paraphrasing methods, sentence extraction and compression, sentence and surface realizers, and text summarization [3, 5, 15, 7]. As a future challenge for more attractive advertisements, the system could take into account sentiment analysis as well. Further steps of our future work regarding expansion of user requirements from our application are: a. Improving task scheduler and take into consideration the possible concurrent use of the system from a large volume of new users. b. Allowing users to select if they want to construct a completely new

campaign or to optimize an already running one, regarding bidding adjustments of existing keywords. For the latter, our system will migrate the information that is being kept in our database tables to the existing settings of the AdWords Campaign.

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

References

1. Abhishek, V., Hosanagar, K.: Keyword generation for search engine advertising using semantic similarity between terms. In: ICEC, pp. 89–94 (2007)
2. Abhishek, V., Hosanagar, K.: Optimal bidding in multi-item multi-slot sponsored search auctions. In: EC '12 (2012)
3. Androutsopoulos, I., Malakasiotis, P.: A survey of paraphrasing and textual entailment methods. J. Artif. Intell. Res. (JAIR) **38**, 135–187 (2010)
4. Borgs, C., Chayes, J., Immorlica, N., Jain, K., Etesami, O., Mahdian, M.: Dynamics of bid optimization in online advertisement auctions. WWW '07, pp. 531–540 (2007)
5. Choi, Y., Fontoura, M., Gabrilovich, E., Josifovski, V., Mediano, M., Pang, B.: Using landing pages for sponsored search ad selection. In: Proceedings of the 19th international conference on World wide web, WWW '10, pp. 251–260. ACM, New York, NY, USA (2010)
6. Gabrilovich, E.: Ad retrieval systems *in vitro* and *in vivo*: Knowledge-based approaches to computational advertising. In: ECIR, pp. 4–5 (2011)
7. Galanis, D., Androutsopoulos, I.: An extractive supervised two-stage method for sentence compression. In: HLT-NAACL, pp. 885–893 (2010)
8. Joshi, A., Motwani, R.: Keyword generation for search engine advertising. In: ICDM Workshops, pp. 490–496 (2006)
9. Larman, C., Basili, V.: Iterative and incremental developments. a brief history. Computer **36**(6), 47 –56 (2003)
10. Liakopoulos, K., Thomaidou, S., Vazirgiannis, M.: The adomaton prototype: Automated online advertising campaign monitoring and optimization. In: 8th Ad Auctions Workshop, EC '12 (2012)
11. Liu, J., Wang, C., Liu, Z., Yao, W.: Advertising keywords extraction from web pages. WISM'10, pp. 336–343 (2010)
12. Ravi, S., Broder, A.Z., Gabrilovich, E., Josifovski, V., Pandey, S., Pang, B.: Automatic generation of bid phrases for online advertising. In: WSDM, pp. 341–350 (2010)
13. Thomaidou, S., Leymonis, K., Liakopoulos, K., Vazirgiannis, M.: Ad-mad: Integrated system for automated development and optimization of online advertising campaigns. In: Demo to appear in ICDM '12 Workshop Proceedings (2012)
14. Thomaidou, S., Vazirgiannis, M.: Multiword keyword recommendation system for online advertising. In: ASONAM 2011, pp. 423–427 (2011)
15. Zhang, Y., Zincir-Heywood, N., Milios, E.: World wide web site summarization. Web Intelli. and Agent Sys. **2**, 39–53 (2004)