Ad Retrieval Systems *in vitro* and *in vivo*: Knowledge-Based Approaches to Computational Advertising

Evgeniy Gabrilovich

Yahoo! Research, California, USA
`gabr@yahoo-inc.com`

Abstract. Over the past decade, online advertising became the principal economic force behind many an Internet service, from major search engines to globe-spanning social networks to blogs. There is often a tension between online advertising and user experience, but on the other hand, advertising revenue enables a myriad of free Web services to the public and fosters a great deal of innovation. Matching the advertisers' message to a receptive and interested audience benefits both sides; indeed, literally hundreds of millions of users occasionally click on the ads, hence they should be considered relevant to the users' information needs by current IR evaluation principles. The utility of ads can be better explained by considering advertising as a medium of information [2, 3]. Similarly to aggregated search [1], which enhances users' Web search experience with relevant news, local results, user-generated content, or multimedia, online advertising provides another rich source of content. This source, however, is in a complexity class of its own, due to the brevity of bid phrases, ad text being optimized for presentation rather than indexing, and multiple, possibly contradictory utility functions.

A new scientific sub-discipline—Computational Advertising—has recently emerged, which strives to make online advertising integral to the user experience and relevant to the users' information needs, as well as economically worthwhile to the advertiser and the publisher. In this talk we discuss the unique algorithmic challenges posed by searching the ad corpus, and report on empirical evaluation of large-scale advertising systems *in vivo*. At first approximation, finding user-relevant ads is akin to ad hoc information retrieval, where the user context is distilled into a query executed against an index of ads. However, the elaborate structure of ad campaigns, along with the cornucopia of pertinent non-textual information, makes ad retrieval substantially and interestingly different. We show how to adapt standard IR methods for ad retrieval, by developing structure-aware indexing techniques and by augmenting the ad selection process with exogenous knowledge. Computational advertising also employs a host of NLP techniques, from text summarization for just-in-time ad matching, to machine translation for cross-language ad retrieval, to natural language generation for automatic construction of advertising campaigns. Last but not least, we study the interplay between the algorithmic and sponsored search results, as well as formulate and explore context transfer, which characterizes the user's transition from

Web search to the context of the landing page following an ad-click. These studies offer deep insights into how users interact with ads, and facilitate better understanding of the much broader notion of relevance in ad retrieval compared to Web search.

Biography: Evgeniy Gabrilovich is a Senior Research Scientist and Manager of the NLP & IR Group at Yahoo! Research, where he works on advancing the state of the art in information retrieval and natural language processing at Web scale. He has filed over 20 patents, and has over 40 publications in top international venues such as ACM TOIS, JMLR, JAIR, WWW, and WSDM. Evgeniy served as a Senior PC member or Area Chair at SIGIR, AAAI, IJCAI, EMNLP, and ICWSM, and also served on the program committees of virtually every major conference in the field.

Evgeniy earned his MSc and PhD degrees in Computer Science from the Technion—Israel Institute of Technology. In his PhD thesis, he developed a methodology for using large scale repositories of world knowledge (e.g., all the knowledge available in Wikipedia) to enhance text representation beyond the bag of words. Subsequently, he also worked on combining heterogenous knowledge sources, such as Web search results, classification with respect to an external taxonomy, and users' actions, for improving the accuracy of information retrieval. Evgeniy researched how to interpret information in context, and devised methods for analyzing the context of user queries. He developed one of the most commonly used datasets for computing semantic relatedness of words, as well as formulated a methodology for parameterized generation of labeled datasets for text categorization. Evgeniy taught multiple tutorials on computational advertising and organized a number of workshops at top conferences, including a workshop on feature generation and selection for information retrieval at SIGIR, and workshops on the synergy between user-contributed knowledge and research in AI at IJCAI and AAAI.

References

1. Jaime Arguello, Fernando Diaz, Jamie Callan, and Jean-Francois Crespo. Sources of evidence for vertical selection. In *Proceedings of the 31st Annual International ACM SIGIR Conference*, pages 315–322, 2009.
2. Phillip Nelson. Advertising as information. *Journal of Political Economy*, 82(4):729–754, 1974.
3. David Ogilvy. *Ogilvy on Advertising*. Vintage Books, 1985.