

Information Retrieval Challenges in Computational Advertising

Andrei Broder, Evgeniy Gabrilovich, Vanja Josifovski

Yahoo! Research, Santa Clara, CA, USA
{broder, gabr, vanjaj} @ yahoo-inc.com

Abstract:

Computational advertising is an emerging scientific sub-discipline, at the intersection of large scale search and text analysis, information retrieval, statistical modeling, machine learning, classification, optimization, and microeconomics. The central challenge of computational advertising is to find the "best match" between a given user in a given context and a suitable advertisement. The aim of this tutorial is to present the state of the art in Computational Advertising, in particular in its IR-related aspects, and to expose the participants to the current research challenges in this field. The tutorial does not assume any prior knowledge of Web advertising, and will begin with a comprehensive background survey.

Going deeper, our focus will be on using a textual representation of the user context to retrieve relevant ads. At first approximation, this process can be reduced to a conventional setup by constructing a query that describes the user context and executing the query against a large inverted index of ads. We show how to augment this approach using query expansion and text classification techniques tuned for the ad-retrieval problem. In particular, we show how to use the Web as a repository of query-specific knowledge and use the Web search results retrieved by the query as a form of a relevance feedback and query expansion. We also present solutions that go beyond the conventional bag of words indexing by constructing additional features using a large external taxonomy and a lexicon of named entities obtained by analyzing the entire Web as a corpus.

The last part of the tutorial will be devoted to a potpourri of recent research results and open problems inspired by Computational Advertising challenges in text summarization, natural language generation, named entity recognition, computer-human interaction, and other SIGIR-relevant areas.

ACM Categories & Descriptors:

H.3.3 Information Search and Retrieval

General Terms:

Algorithms, Experimentation.

Keywords:

Online advertising, sponsored search, content match

Bio/Bios:

Andrei Broder is a Yahoo! Fellow and Vice President Computational Advertising. Previously he was an IBM Distinguished Engineer and the CTO of the Institute for Search

and Text Analysis in IBM Research. From 1999 until 2002 he was VP for Research and Chief Scientist at the AltaVista Company. He was graduated Summa cum Laude from Technion, the Israeli Institute of Technology, and obtained his M.Sc. and Ph.D. at Stanford University under Don Knuth. He has authored more than a hundred papers and was awarded twenty-five patents. He is a member of the US National Academy of Engineering, a fellow of ACM and of IEEE, and past chair of the IEEE Technical Committee on Mathematical Foundations of Computing.

http://research.yahoo.com/Andrei_Broder

Evgeniy Gabrilovich is a Senior Research Scientist and Manager of the NLP & IR Group at Yahoo! Research. Recently, he organized a workshop on the synergy between user-contributed knowledge and research in AI at IJCAI'09, and a workshop on information retrieval for advertising at SIGIR'09. Evgeniy presented tutorials on computational advertising at CIKM'09, IJCAI'09, ACL'08, and EC'08. He served on the program committees of WWW, WSDM, SIGIR, CIKM, AAAI, ACL, EMNLP, ICWSM, HLT, COLING, and JCDL. Evgeniy earned his M.Sc. and Ph.D. degrees in Computer Science from the Technion - Israel Institute of Technology.

http://research.yahoo.com/Evgeniy_Gabrilovich

Vanja Josifovski is a Principal Research Scientist at Yahoo! Research, where he works on search and advertisement technologies for the Internet. He is currently exploring designs for the next generation ad placement platforms for textual and behavioral advertising. Previously, Vanja was a Research Staff Member at the IBM Almaden Research Center working on several projects in database runtime and optimization, federated databases, and enterprise. He earned his M.Sc. degree from the University of Florida at Gainesville and his Ph.D. from the Linköping University in Sweden. Vanja has published over 50 peer reviewed publications, authored around 40 patent applications, and was on the program committees of WWW, SIGIR, ICDE, VLDB, CIKM, ICDM, KDD and other major conferences in the database, information retrieval, and search areas.

http://research.yahoo.com/Vanja_Josifovski