Contents

Chapter 1 Introduction - Project Context **4**
1.1 Online Advertising 4
1.2 Computational Advertising 4

Chapter 2 Project Objectives and Specifications **5**
2.1 Problem specification 5
2.2 Functional requirements 5
2.3 Non-Functional requirements 5

Chapter 3 Bibliographic research **6**
3.1 Keyword Extraction 6
3.1.1 Candidate Selection 6
3.1.2 Keyword Classification 6
3.2 Topic Modeling 6
3.2.1 Topic Models 6
3.2.2 The latent topic structure of documents 6
3.2.3 Similarity computation at topic level 6
3.2.4 Recommending high topic coverage 6
3.3 User history 6
3.3.1 Dynamic user interests 6

Chapter 4 Analysis and Theoretical Foundation **7**
4.1 Overall Architecture Analysis 8
4.2 Content Preprocessing 8
4.2.1 Context parsing 8
4.2.2 Stop word removal and Stemming 8
4.3 Keyword Extraction 8
4.3.1 Data Aquisition 8
4.3.2 Feature Selection 8
4.3.3 Classifier Choice 8
4.4 Topic Mapping 8
4.4.1 Topic Model Description 8

4.4.2 Unified Topic Representation 8
4.5 Coverage Based Recommendation 8
4.5.1 Contextual Recommendation 8
4.5.2 Behavioral Recommendation 8
4.5.3 Maximizing Topic Model Coverage 8

Chapter 5 Detailed Design and Implementation **9**
5.1 System architecture 9
5.1.1 Online flow 9
5.1.2 Offline flow 9
5.2 Topic extraction process 9
5.2.1 Advertisement topic identification 9
5.2.2 Context topics mapping 9
5.2.3 User topic trends 9
5.3 Coverage Based Recommendation 9
5.4 Implementation details 9

Chapter 6 Testing and Validation **10**
6.1 Technological motivation 10
6.2 Compared Relevance 10
6.3 Coverage Coefficient 10
6.4 System Scalability 10

Chapter 7 User's manual **11**
7.1 System Installation 11
7.2 User's Manual 11

Chapter 8 Conclusions **12**
8.1 Contributions and Achievements 12
8.2 Result Analysis 12
8.3 Further Work 12

Bibliography **13**

Bibliography **13**

Bibliography

[1] C. D. Manning, P. Raghavan, and H. Schtze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.

[2] H. Garcia-Molina, G. Koutrika, and A. Parameswaran, "Information seeking: convergence of search, recommendations, and advertising," *Commun. ACM*, vol. 54, no. 11, pp. 121–130, Nov. 2011. [Online]. Available: http://doi.acm.org/10.1145/2018396.2018423

[3] A. Ahmed, Y. Low, M. Aly, V. Josifovski, and A. J. Smola, "Scalable distributed inference of dynamic user interests for behavioral targeting," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 114–122. [Online]. Available: http://doi.acm.org/10.1145/2020408.2020433

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944937

[5] A. Broder, M. Fontoura, V. Josifovski, and L. Riedel, "A semantic approach to contextual advertising," in *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '07. New York, NY, USA: ACM, 2007, pp. 559–566. [Online]. Available: http://doi.acm.org/10.1145/1277741.1277837

[6] B. Zhang, H. Li, Y. Liu, L. Ji, W. Xi, W. Fan, Z. Chen, and W.-Y. Ma, "Improving web search results using affinity graph," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 504–511. [Online]. Available: http://doi.acm.org/10.1145/1076034.1076120

[7] D. M. Blei, "Introduction to probabilistic topic models," *Communications of the ACM*, 2011. [Online]. Available: http://www.cs.princeton.edu/ blei/papers/Blei2011.pdf

[8] Y. Zhang, A. C. Surendran, J. C. Platt, and M. Narasimhan, "Learning from multi-topic web documents for contextual advertisement," in *Proceedings of the 14th*

ACM SIGKDD international conference on Knowledge discovery and data mining, ser. KDD '08. New York, NY, USA: ACM, 2008, pp. 1051–1059. [Online]. Available: http://doi.acm.org/10.1145/1401890.1402015

[9] W.-t. Yih, J. Goodman, and V. R. Carvalho, "Finding advertising keywords on web pages," in *Proceedings of the 15th international conference on World Wide Web*, ser. WWW '06. New York, NY, USA: ACM, 2006, pp. 213–222. [Online]. Available: http://doi.acm.org/10.1145/1135777.1135813

[10] B. Ribeiro-Neto, M. Cristo, P. B. Golgher, and E. Silva de Moura, "Impedance coupling in content-targeted advertising," in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, ser. SIGIR '05. New York, NY, USA: ACM, 2005, pp. 496–503. [Online]. Available: http://doi.acm.org/10.1145/1076034.1076119

[11] G. Heinrich, "Parameter estimation for text analysis," Tech. Rep., 2009.

[12] D. Chakrabarti, D. Agarwal, and V. Josifovski, "Contextual advertising by combining relevance with click feedback," in *Proceedings of the 17th international conference on World Wide Web*, ser. WWW '08. New York, NY, USA: ACM, 2008, pp. 417–426. [Online]. Available: http://doi.acm.org/10.1145/1367497.1367554

[13] A. Broder and V. Josifovki, "Introduction to computational advertising - lecture 4: Contextual advertising," MSE 239, Stanford University, California, September-December 2011. [Online]. Available: http://www.stanford.edu/class/msande239

[14] D. M. Blei, "Advanced methods in probabilistic modeling - lecture: Mixed membership models," COS597C, Princeton University, New Jersey, Fall 2011. [Online]. Available: http://www.cs.princeton.edu/courses/archive/fall11/cos597C/lectures/mixed-membership.pdf