Improving Semantic Consistency of Web Sites by Quantifying User Intent

Carsten Stolz[1], Maximilian Viermetz[2], Michal Skubacz[3],Ralph Neuneier[4]

[1] University of Eichstätt-Ingolstadt, Germany, carsten.stolz@ku-eichstaett.de
[2] Heinrich-Heine-Universität Düsseldorf, Germany, maximilian@viermetz.net
[3] Siemens AG, Corporate Technology, Germany, michal.skubacz@siemens.com
[4] Siemens Corporate Research Princeton, USA, ralph.neuneier@siemens.com

Abstract. The design and organization of a website reflects the authors intent. Since user perception and understanding of websites may differ from the authors, we propose a means to identify and quantify this difference in perception. In our approach we extract perceived semantic focus by analyzing user behavior in conjunction with keyword similarity.
By combining usage and content data we identify user groups with regard to the subject of the pages they visited. Our real world data shows that these user groups are nicely distinguishable by their content focus. By introducing a distance measure of keyword coincidence between web pages and user groups, we can identify pages of similar perceived interest. A discrepancy between perceived distance and link distance in the web graph indicates an inconsistency in the web site's design. Determining usage similarity allows the web site author to optimize the content to the users needs. [5]

References

1. Berendt, B.; Hotho, A.; Stumme, G., Towards Semantic Web Mining, The Semantic Web - ISWC 2002, 2002, p. 264
2. Calero, C.; Ruiz, J.; Piattini, M., A Web Metrics Survey Using WQM Web Engineering, 4th Int. Conf. ICWE 2004, Proc. p.147-160
3. Cooley,R. The Use of Web Structure and Content to Identify Subjectively Interesting Web Usage Patterns ACM Transaction on Internet Technology, 2003, Vol.3 Nr.2 p. 93–116
4. Dai,H.,Mobasher,B., Using ontologies to discover domain-level web usage profiles,PKDD 2001
5. Dhyani, D.;Keong NG, W.;Bhowmick, S.S., A Survey of Web Metrics, ACM Computing Surveys, 2002, vol. 34, nr. 4, p. 469-503
6. B. Mobasher, H. Dai, T. Luo, Y. Sun, and J. Zhu Integrating Web Usage and Content Mining for More Effective Personalization. Proc. of the Int'l Conf. on E-Commerce and Web Technologies (ECWeb2000)(2000)

7. D. Oberle, B. Berendt, A. Hotho, J. Gonzalez, Conceptual User Tracking,Proceedings of the Atlantic Web Intelligence Conference,2002, p. 155 - 164
8. Stolz,C., Gedov,V., Yu,K., Neuneier,R., Skubacz,M. Measuring Semantic Relations of Web Sites by Clustering of Local Context, ICWE2004, München (2004) In Proc. Int. Conf. on Web Engineering 2004, Springer, p. 182–186
9. M. F. Porter. An algorithm for suffix stripping. Program, 14:130–137, (1980)
10. Vadnerdonckt, J.; Beirekdar, A.; Noirhomme-Fraiture, M.; Automated Evaluation of Web Usability and Accessibility by Guideline Review, In Proc. Int. Conf. Web Engineering 2004, Springer, p. 17-30
11. Zhu, J.;Hong, J.; Hughes, J.G., PageCluster: Mining Conceptual Link Hierarchies from Web Log Files for Adaptive Web Site Navigation, ACM Journal Transaction on Internet Technology,2004, Vol.4,Nr.2, p. 185-208
12. A.Sun and E.-P. Lim. Web Unit Mining: Finding and Classifying Subgraphs of Web Pages. In Proceedings 12th Int. Conf. on Information and Knowledge Management, p. 108–115. ACM Press, 2003.
13. S. Chakrabarti; Mining the Web - Discovering Knowledge from Hypertext Data. Morgan Kaufmann, 2002
14. X. He, H. Zha, C. Ding, and H. Simon Web document clustering using hyperlink structures Computational Statistics and Data Analysis, 41:19-45, 2002
15. R. Song and H. Liu and J. Wen and W. Ma Learning important models for web page blocks based on layout and content analysis SIGKDD Explor. Newsl. vol.6, nr. 2, 2004, p.14–23