WebKDD 2006 – Web Mining and Web Usage Analysis Post-Workshop report

Olfa Nasraoui
Computer Engineering and Computer Science department
Speed School of Engineering
University of Louisville
Louisville KY 40292
olfa.nasraoui@louisville.edu

Myra Spiliopoulou
Otto-von-Guericke-University Magdeburg,
Faculty of Computer Science
D-39016 Magdeburg, Germany
myra@iti.cs.uni-magdeburg.de

Jaideep Srivastava
Dept. of Computer Science and Engineering
University of Minnesota
Minneapolis, MN 55455
srivasta@cs.umn.edu

Bamshad Mobasher
School of Computer Science, Telecommunications & Information Systems, DePaul University
Chicago, IL 60604
mobasher@cti.depaul.edu

Brij Masand
Data Miners, Inc
77 North Washington Street, 2nd Floor
Boston , MA 02114
brij@data-miners.com

ABSTRACT

In this report, we summarize the contents and outcomes of the recent WebKDD 2006 workshop on Web Mining and Web Usage Analysis that was held in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23, 2006, in Philadelphia, Pennsylvania. In 2006, WebKDD was organized for the eighth time. It solicited papers on the broad overview subject "*Knowledge Discovery on the Web*" and on new directions for Web mining research. We reflect on the results of the workshop, as captured in presentations and plenary discussions.

Keywords

Web mining, Web analytics, personalization, recommender systems, collaborative filtering, spam detection, blog mining, search engine query log mining, peer to peer personalization

1. INTRODUCTION

WebKDD 2006 is the eigth of a successful series of workshops on knowledge discovery on the Web. The WebKDD '06 workshop continued its tradition of serving as a bridge between academia and industry by bringing together practitioners and researchers from both areas in order to foster the exchange of ideas and the dissemination of emerging solutions for intelligent Web-based applications using Web usage, structure and content mining.

2. THEME

The theme of the 8th WebKDD workshop was "Knowledge Discovery on the Web", encompassing lessons learned over the past years and new challenges for the years to come. While solutions to some of the infancy problems of Web analysis have reached maturity, the reality poses new challenges: The Web is evolving constantly; sites change and user preferences drift. And, most of all, a Web site is more than a see-and-click medium; it is a venue where a user interacts with a site owner or with other users, where group behavior is exhibited, communities are formed and experiences are shared. The WebKDD '2006 workshop invited research results in all areas of Web mining and Semantic Web mining, with an emphasis on a seven years' update: What are the lessons learned on algorithms, semantics, data preparation, data integration and applications of the Web? How are new technologies, like adaptive mining methods, stream mining algorithms and techniques for the Grid apply to Web mining? What new challenges are posed by new forms of data, especially flat texts, documents, pictures and streams, as well as the emergence of Web communities? How do we study the evolution of the Web and its effects on searching and browsing behavior? Which lessons have we learned about usability, e-commerce applications, personalization, recommendation engines, Web marketplaces, Web search, Web security, and misuse and abuse of the Web and its services?

3. SUBMISSIONS

24 papers were submitted to WebKDD 2006 and were strictly reviewed by three reviewers from the WebKDD 2006 program committee. Due to the quality and diversity of the submissions, 10 papers were accepted for full presentations and 7 for short presentations (see the References section). The workshop papers, which are available at http://webmining.spd.louisville.edu/WebKDD06/proceedings.pdf, have been divided into four categories depending on their topics: (i) Adaptive Web Navigation & Collaborative Filtering, (ii) Web Content, Search, and Queries, (iii) Recommender Systems, (iv) Web Mining Emerging Trends.

4. WORKSHOP

The workshop attracted interest from a large number of conference participants. Over 140 conference attendees had pre-registered for WebKDD '06 before it started, and at any given time throughout the day, there were between 40 and 60 participants attending the talks. In addition to participants from academia and government agencies, a significant proportion of the workshop participants came from the Web related industry, including Amazon, Yahoo, Google, Microsoft, AOL, Ask.com, E-Bay, CNET, and a few other companies. The paper presentations were divided into four sessions according to their main focus, as described below.

4.1 Session 1: Adaptive Web Navigation & Collaborative Filtering

The first paper, "*Adaptive Website Design using Caching Algorithms*" by Justin Brickell, Inderjit S. Dhillon, and Dharmendra S. Modha, presents improved online algorithms for shortcut link selection that are based on a novel analogy drawn between shortcutting and caching. In the same way that cache algorithms predict which memory pages will be accessed in the future, the proposed algorithms predict which web pages will be accessed in the future. These algorithms are efficient and can consider accesses over a long period of time, but give extra weight to recent accesses. Experiments show significant improvement in the utility of shortcut links selected by the proposed algorithm as compared to those selected by existing algorithms.

The second paper, "*Usage-Aware Average Clicks*" by Kalyan Beemanapalli, Ramya Rangarajan, and Jaideep Srivastava, present an extension to the Average-Clicks Algorithm, called Usage Aware Average-Clicks, where the static web link structure graph is combined with the dynamic Usage Graph (built using the information available from the web logs) to assign different weights to links on a web page and hence capture the user's intuition of distance between two Web pages more accurately. This method has been used as a new metric to calculate the page similarities in a recommendation engine to improve its predictive power.

The third paper, "*Nearest-Biclusters Collaborative Filtering*" by Panagiotis Symeonidis, Alexandros Nanopoulos, Apostolos Papadopolos, and Yannis Manolopoulos, uses biclustering to disclose the duality between users and items in Nearest-neighbor Collaborative Filtering, by grouping them in both dimensions simultaneously. A novel nearest-biclusters algorithm is proposed, that uses a new similarity measure that achieves partial matching of users' preferences. Performance evaluation results are offered, which show that the proposed method improves substantially the performance of the CF process.

4.2 Session 2: Web Content, Search, and Queries

The first paper, "*Model-Based Classification of Web Documents Represented by Graphs*" by Alex Markov, Mark Last, and Abraham Kandel, addresses the limitations of the vector-space model of information retrieval does not capture important structural information, such as the order and proximity of word occurrence, the location of a word within the document, or mark-up information. Three new hybrid approaches to web document classification are presented, built upon both graph and vector space representations, thus preserving the benefits and discarding the limitations of each. The hybrid methods outperform, in most cases vector-based models using two model-based classifiers (C4.5 decision-tree algorithm and probabilistic Naïve Bayes) on several benchmark web document collections.

The second paper, "*Keyword Weight Propagation for Indexing Structured Web Content*" by Jong Wook Kim and K. Selcuk Candan, addresses the problem of indexing Web documents having non-atomic structures, such as navigational/semantic hierarchies on the Web. A novel keyword and keyword weight propagation technique are proposed to properly enrich the data nodes in structured content. The approach first relies on understanding the context provided by the relative content relationships between entries in the structure, and then leveraging this information for relative-content preserving keyword propagation. Experiments show a significant improvement (10−15%) in precision with the proposed keyword propagation algorithm

The third paper, "*How to Define Searching Sessions on Web Search Engines*" by Bernard J. Jansen, Amanda Spink, Vinish Kathuria, and Sherry Koshman, investigates three methods for defining a session on Web search engines. The authors examine 2,465,145 interactions from 534,507 Web searchers, and compare defining sessions using: 1) Internet Protocol address and cookie; 2) Internet Protocol address, cookie, and a temporal limit on intra-session interactions; and 3) Internet Protocol address, cookie, and query reformulation patterns. Research results shows that defining sessions by query reformulation along with Internet Protocol address and cookie, provides the best measure, resulting in an 82% increase in the number of sessions; while for all methods, mean session length was fewer than three queries and the mean session duration was less than 30 minutes. Implications are that unique sessions may be a better indicator than the common industry metric of unique visitors for measuring search traffic.

4.3 Session 3: Recommender Systems

The first paper, "*Incorporating Concept Hierarchies into Usage Mining Based Recommendations*" by Amit Bose, Kalyan Beemanapalli, Jaideep Srivastava, and Sigal Sahar, addresses the limitation of most recommendation models in their ability to use domain knowledge in the form of conceptual and structural characteristics of a website. Conceptual content organization can play an important role in the quality of recommendations, and forms the basis of resources like Google Directory, Yahoo Directory and web-content management systems. The authors propose a novel technique to incorporate the conceptual characteristics of a website into a usage-based recommendation model. The authors use a framework based on biological sequence alignment. Similarity scores play a crucial role in such a construction, and a scoring system that is generated from the

website's concept hierarchy is introduced. These scores fit seamlessly with other quantities used in similarity calculation like browsing order and time spent on a page. Additionally they demonstrate a simple, extensible system for assimilating more domain knowledge. Experimental results illustrate the benefits of using a concept hierarchy.

The second paper, "*A Random-Walk Based Scoring Algorithm with Application to Recommender Systems for Large-Scale E-Commerce*" by Augusto Pucci and Marco Gori, presents "ItemRank", a random-walk based scoring algorithm, which can be used to rank products according to expected user preferences, in order to recommend top-rank items to potentially interested users. The authors tested their algorithm on the MovieLens data set, which contains data collected from a popular recommender system on movies, and compared ItemRank with other state-of-the-art ranking techniques, showing that ItemRank performs better than the other techniques, while being less complex than other algorithms with respect to memory usage and computational cost. The paper also presents an analysis that helps to discover some intriguing properties of the MovieLens data set, that has been widely exploited as a benchmark for evaluating recently proposed approaches to recommender system.

The third paper in this category, "*ClustKNN: A Highly Scalable Hybrid Model- & Memory-Based CF Algorithm*" by Al Mamunur Rashid, Shyong K. Lam, George Karypis, and John Riedl, addresses the need for specially designed CF algorithms that can gracefully cope with the vast size of the data representing customers and items in typical e-commerce systems. Many algorithms proposed thus far, where the principal concern is recommendation quality, may be too expensive to operate in a large-scale system. The authors propose ClustKNN, a simple and intuitive algorithm that is well suited for large data sets. The method first compresses data tremendously by building a straightforward but efficient clustering model. Recommendations are then generated quickly by using a simple Nearest Neighbor-based approach. The feasibility of ClustKNN is demonstrated both analytically and empirically, and a comparison with a number of other popular CF algorithms shows that, apart from being highly scalable and intuitive, ClustKNN provides very good recommendation accuracy.

The fourth paper, "*Recommendation Based on Influence Sets*" by Jian Chen and Jian Yin, addresses the sensitivity to sparsity suffered by traditional item-based collaborative filtering (CF) algorithms, by presenting a different view of integrating Influence Sets into the recommendation process. Recommendation Based on Influence Sets (RIS) is a novel item-based CF approach that combines the effects of K nearest neighbors with reverse K nearest neighbors for a target item to enhance the density of information. New prediction generation methods are defined for this new recommendation mechanism, and experimental results show that RIS can achieve better prediction accuracies than traditional item-based CF algorithm and alleviate the dataset sparsity problem.

The last paper in this category, "*Relevance of Time Spent on Web Pages*" by Peter I. Hofgesang, tries to advance the state of the art in answering one of the most important questions in Web personalization: What is the intention of an online visitor? Online, the users' intentions are hidden and can only be partially revealed from implicit indicators within the traces left behind while they browse through a website. While the vast majority of research in web usage mining exploit only two types of information: the ordering of visited pages and their popularity - the number of times they were visited, a third factor, the *time spent on web pages* (TSP) has been suggested by several studies in information retrieval and human-computer interaction as an important measure of intention and relevance. Thus, the authors present (1) an extensive survey over possible factors that influence the TSP measure and (2) a similarity measure that applies to the TSP and can be used to cluster users based on their assumed intentions. Experiments are based on log files generated by several commercial web-sites.

4.4 Session 4: Web Mining Emerging Trends

The first paper, "*The Impact of Attack Profile Classification on the Robustness of Collaborative Recommendation*" by Chad Williams, Runa Bhaumik, Robin Burke, and Bamshad Mobasher, addresses the vulnerability of Collaborative recommender systems to profile injection attacks. By injecting a larger number of biased profiles into a system, attackers can manipulate the predictions of targeted item. To decrease this risk, researchers have begun to study mechanisms for detecting and preventing profile injection attacks. In this paper, the authors extend their previous work that proposed several attributes for attack detection and for classification of attack profiles, through a more detailed analysis of the informativeness of these attributes as well as an evaluation of their impact at improving the robustness of recommender systems.

The second paper, "*Client-side Web Mining for Community Formation in Peer-to-Peer Environments*" by Kun Liu, Kanishka Bhaduri, Kamalika Das, Phuong Nguyen, and Hillol Kargupta, presents a framework for self-formation of interests-based Peer-to-Peer communities using client-side web browsing history. The authors propose an order statistics-based algorithm to build communities with hierarchical structures, and also carefully consider the privacy concerns of the peer by adopting cryptographic protocols to measure similarity between peers without disclosing their personal profiles. The algorithm is evaluated using a distributed data mining toolkit, and experimental results show that the proposed framework could efficiently build interests-based communities.

The third paper, "*Using Rank Propagation and Probabilistic Counting for Link-Based Spam Detection*" by Luca Becchetti, Carlos Castillo, Debora Donato, Stefano Leonardi, and Ricardo Baeza-Yates, describes a technique for automating the detection of Web link spam, that is, groups of pages that are linked together with the sole purpose of obtaining an undeservedly high score in search engines. The

problem of Web spam is widespread and difficult to solve, mostly due to the large size of Web collections that makes many algorithms unfeasible in practice. For spam detection the authors apply only link-based methods, that consider the topology of the Web graph without looking at the contents of the pages. Web page attributes are computed by applying rank propagation and probabilistic counting over the Web graph, and then used to build a classifier, based only on link data, that detected about 80% of the spam pages with a rate of false positives of 2%. This result is competitive with state-of-the-art spam classifiers that use content attributes.

The last paper in this category, "Mining Sentiment Classification from Political Web Logs" by Kathleen T. Durant and Michael D. Smith, investigates existing technology and their utility for sentiment classification on web log posts that allow people to share their opinions on a wide range of "hot" topics with virtual communities. The authors show that a Naïve Bayes classifier can on average correctly predict a posting's political category 78.06% of the time, and significantly outperforms Support Vector Machines (SVMs) which in turn correctly predicted the category of web log posts 75.47% of the time. Previous research was able to achieve an 81.0% accuracy using Naïve Bayes and 82.9% using SVMs using the chosen feature set representation on a nonspecific topic corpus. Using a dataset of political web logs over a two-year period, they also show that it is important to maintain a uniform distribution in such datasets to avoid biases in classification.

5. CONCLUSIONS AND FUTURE DIRECTIONS

The WebKDD '06 workshop continued a tradition of exposing strong and ever more innovative research in critical areas of Web usage mining and personalization, in particular, collaborative filtering and recommender systems in general [1,2,3,9,10,11,12,13,14]. Moreover, this year saw increasing interest in new topics, mostly invigorated by the popularity of search engines and the new potential of the Web 2.0. Such topics included data mining on search engine data, e.g. query logs [6,7,8], richer mining of content [4,5,7], as well as Web mining in peer to peer environments [15], link based spam detection [16], and mining Web blogs [17]. In addition to the discussions that followed each presentation, there was an exciting discussion among all the participants in the workshop just before closing the WebKDD '06 workshop. These discussions helped raise several questions and bring to light several interesting future directions and challenges in the area of Web mining. In particular, participants discussed the importance of research in new problems and applications that are becoming more prevalent in the Web 2.0 context. While some have emphasized the importance of research in adversarial information retrieval, others have pointed to the crucial need and potential for research in mining such sources as semantics tags, peoples' comments on community websites and on blogs, as well as mining social network websites such as MySpace. The need for Web recommender systems that can intelligently explain the basis for their recommendations has also been discussed, as well as the need for more research in contextual information access. Like in previous years at WebKDD , the participants also raised the issue of privacy and the need to make certain data sets available to stimulate research, while not violating privacy. These last two goals seem to be still contradictory; unfortunately, this makes access to data for research purposes a hard task. We hope that the discussions and this report will encourage researchers to explore some of these challenges, and perhaps present some of their answers in future workshops.

6. ACKNOWLEDGMENTS

We would like to thank the authors of all submitted papers. Their creative efforts have supported the competitive technical program of WebKDD 2006. We would also like to express our gratitude to the members of the Program Committee for their vigilant and timely reviews, namely: *Corin Anderson, Ricardo A. Baeza-Yates, Bettina Berendt, Zheng Chen, Ed H. Chi, Brian D. Davison, Wei Fan, Fabio Grandi, Michael Hahsler, Xin Jin, Thorsten Joachims, George Karypis, Ravi Kumar, Vipin Kumar, Mark Last, Mark Levene, Ee-Peng Lim, Huan Liu, Stefano Lonardi, Alexandros D. Nanopoulos, Georgios Paliouras, Aniruddha G. Pant, Jian Pei, Ellen Spertus, Andrew Tomkins,* and *Mohammed J. Zaki.* We extend our thanks and encouragements to the entire Web Mining community, including all those who attended the workshop and contributed in the interesting discussions. Last but not least, we are grateful to *Stefano Lonardi* for allowing us to use a photo that he took (see Figure 1) for the WebKDD '06 brochure and proceedings cover. Olfa Nasraoui gratefully acknowledges the support of *NSF* as part of NSF CAREER award IIS-0133948.

Picture taken by Stefano Lonardi and used for the WebKDD '06 brochure.

Figure 1: resemblance between the light streaks and web clickstreams.

7. REFERENCES

[1] Justin Brickell, Inderjit S. Dhillon, Dharmendra S. Modha, "Adaptive Website Design using Caching Algorithms", In Proc. of *WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis*, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[2] Kalyan Beemanapalli, Ramya Rangarajan, Jaideep Srivastava, "Usage-Aware Average Clicks", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[3] Panagiotis Symeonidis, Alexandros Nanopoulos, Apostolos Papadopolos, and Yannis Manolopoulos,

"Nearest-Biclusters Collaborative Filtering", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[4] Alex Markov, Mark Last, and Abraham Kandel, "Model-Based Classification of Web Documents Represented by Graphs", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[5] Jong Wook Kim and K. Selcuk Candan, "Keyword Weight Propagation for Indexing Structured Web Content", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[6] Bernard J. Jansen, Amanda Spink, Vinish Kathuria, Sherry Koshman, "How to Define Searching Sessions on Web Search Engines", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[7] Ricardo Baeza-Yates, Alvaro Pereira Jr and Nivio Ziviani, "The Evolution of Web Content and Search Engines", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[8] David Nettleton, Liliana Calderon, Ricardo Baeza-Yates, "Analysis of Web Search Engine Query Sessions", In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.

[9] Amit Bose, Kalyan Beemanapalli, Jaideep Srivastava, Sigal Sahar, "Incorporating Concept Hierarchies into Usage Mining Based Recommendations", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[10] Augusto Pucci and Marco Gori, "A Random-Walk Based Scoring Algorithm with Application to Recommender Systems for Large-Scale E-Commerce", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[11] Al Mamunur Rashid, Shyong K. Lam, George Karypis, and John Riedl, "ClustKNN: A Highly Scalable Hybrid Model- & Memory-Based CF Algorithm", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[12] Jian Chen and Jian Yin, "Recommendation Based on Influence Sets", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[13] Peter I. Hofgesang, "Relevance of Time Spent on Web Pages", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[14] Chad Williams, Runa Bhaumik, Robin Burke, and Bamshad Mobasher, "The Impact of Attack Profile Classification on the Robustness of Collaborative Recommendation", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[15] Kun Liu, Kanishka Bhaduri, Kamalika Das, Phuong Nguyen and Hillol Kargupta, "Client-side Web Mining for Community Formation in Peer-to-Peer Environments", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[16] Luca Becchetti, Carlos Castillo, Debora Donato, Stefano Leonardi and Ricardo Baeza-Yates, "Using Rank Propagation and Probabilistic Counting for Link-Based Spam Detection", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

[17] Kathleen T. Durant and Michael D. Smith, "Mining Sentiment Classification from Political Web Logs", *In Proc. of WebKDD 2006: KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), August 20-23 2006, Philadelphia, PA.*

About the authors:

Olfa Nasraoui is the Endowed Chair of E-commerce and the Director of the Knowledge Discovery and Web Mining Lab at the University of Louisville, where she is also Assistant professor in Computer Science and Engineering. She received her Ph.D. in Computer Engineering and Computer Science from the University of Missouri-Columbia in 1999. From 2000 to 2005, she was an Assistant Professor at the University of Memphis. Her research activities include Data Mining, in particular, Web mining and mining evolving data streams, Personalization, and Computational Intelligence.

She has served on the organizing and program committees of several conferences and workshops, including WebKDD 2004 and WebKDD 2005. She is the recipient of the National Science Foundation CAREER Award for outstanding young scientists. Her research is funded mainly by NSF and by NASA. She is a member of IEEE and ACM.
(http:/www.louisville.edu/~o0nasr01)

Myra Spiliopoulou is Professor at the Faculty of Computer Science,Otto-von-Guericke-University Magdeburg,Germany. Her research is on the development and enhancement of knowledge discovery methods for person-web interaction, for text analysis and for knowledge management. Her research on web usage mining and data preparation, text mining, temporal mining and pattern evolution has appeared in journals, conferences, books and workshops. She is regular reviewer in many data mining conferences, including KDD, ECML/PKDD and SIAM Data Mining, of the IEEE TKDE Journal and of many workshops in subjects associated to web mining. She has co-organized most of the WEBKDD workshops in the KDD conference series since 1999, In Europe, she has launched the Web Mining Forum initiative under the KDNet Network of Excellence and organized the workshop "European Web Mining Forum" in ECML/PKDD'2003 and '2005. In the ECML/PKDD conference series she has also given several tutorials on web mining since 1999. In March 2005, she organized the 29th Annual Conference of the German Classification Society (GfKl'2005) in Magdeburg. and in 2006 she was PC Co-Chair of the ECML/PKDD that took place in Berlin (Sept. 2006)
(http://omen.cs.uni-magdeburg.de/itikmd)

Jaideep Srivastava is a Professor of Computer Science & Engineering at the University of Minnesota. He has established and led a laboratory that has conducted research in databases, multimedia systems, and data mining. He has supervised 23 Ph.D. dissertations and 44 MS theses, and has authored/co-authored over 185 papers in journals and conferences. Dr. Srivastava has an active collaboration with the technology industry, both for research and technology transfer, and is an often-invited participant in technical and technology strategy forums. Dr. Srivastava's industry experience includes data mining at Amazon.com and at Yodlee. He has provided technology and technology strategy advice to a number of large corporations, including Cargill, United Technologies, IBM, Honeywell, 3M, and Persistent Systems. He is on the editorial board of a number of journals, and has served as Program and Conference Chair for a number of prominent conferences. Dr. Srivastava has recently been appointed as the Senior Technology Advisor for the State of Minnesota, where he provides advice on information technologies to the State Chief Information Officer. He has been elected a Fellow of the IEEE, and has been appointed a Distinguished Visitor by the IEEE Computer Society. He has a Ph.D. from the University of California, Berkeley, and bachelors from IIT Kanpur, India.
(http://www.cs.umn.edu/people/faculty.php?user=srivasta)

Bamshad Mobasher is an Associate professor of Computer Science and the director of the Center for Web Intelligence (CWI) at DePaul University. He received his PhD from Iowa State University in 1994. His research areas include data mining, Web mining, intelligent agents, and computational logic. He has published more than 70 scientific articles in these areas. As the director of the CWI, Dr. Mobasher directs research in Web mining, Web analytics, user modeling, and personalization; and he oversees several NSF or industry funded projects. Dr. Mobasher has served as an organizer and on the program committees of numerous conferences and workshops, including, the recently held WebKDD workshop on Web Mining and Web Usage Analysis at the 2005 ACM SIGKDD conference in Seattle. He was the local arrangements chair for the 2005 ACM SIGKDD conference (KDD'05) held in Chicago during August 2005.
(http://maya.cs.depaul.edu/~mobasher)

Brij Masand is a partner at Data Miners's Inc. He was formerly head of the data mining group at GTE Labs, where he pioneered web usage mining for analyzing behavior of on-line yellow pages users. He has more than 15 years of experience in applying machine learning technologies to data mining, web usage mining, text mining and intelligent agents and has published numerous papers on these subjects. He has also done extensive work in implementing reliable web usage metrics and applying survival analysis techniques for business applications such as modeling churn and other time-to-event predictions. Brij has an MS in EECS from MIT.
(http://www.data-miners.com/brij/welcome.html)