An overview of web mining in societal benefit areas

Article *in* Online Information Review · April 2008

DOI: 10.1108/14684520810879818 · Source: DBLP

1 author:

Georgios Lappas
Technological Educational Institute of Weste…

An Overview of Web Mining in Societal Benefit Areas

Georgios Lappas
Technological Educational Institution of Western Macedonia
P.O. Box 30, GR-52100 Kastoria, Greece.
lappas@kastoria.teikoz.gr

Abstract

A survey of web mining research related to areas of social benefit is the focus of this paper. We particularly focus on web mining that may help and benefit societal areas in ways of extracting new knowledge, providing support for decision making and empowering valuable management of societal issues. E-commerce and E-business are two fields empowered by web mining having lots of applications for increasing sales and doing intelligence business. Have areas of social interest also been empowered by web mining applications? What is currently the ongoing research and current trends in E-Services areas in fields like E-learning, E-Government, E-Politics and E-democracy? What other areas of social interest can be benefited from web mining? This paper will try to provide the answers by reviewing the literature for applications and methods applied to the above fields.

1. Introduction

The web is constantly becoming a central part of social, cultural, political, educational, academic, and commercial life and contains a wide range of information and applications in areas that are of societal interest. The word 'mining' means extracting something useful or valuable, such as mining gold from the earth. The expectation of useful or valuable information discovery from the web is enclosed in the term "web mining". Definitional, web mining refers to the application of data mining techniques to the World Wide Web [18], or else is the area of data mining that refers to the use of algorithms for extracting patterns from resources distributed in the web. Over the years, web mining has been extended to denote the use of data mining and other similar techniques to discover resources, patterns and knowledge from the web and web-related data [14]. Web mining is a relatively new area, broadly interdisciplinary, attracting researchers from: computer science fields like artificial intelligence, machine learning, databases, and information retrieval specialists; from business studies fields like marketing, administrative and e-commerce specialists; and from social and communication studies fields such as social network analyzers, pedagogical scientists, and political science specialists. Herrera-Viedma and Pasi [31] denote that due to the complexity of web research there is a requirement for the use of interdisciplinary approaches like statistics, databases, information retrieval, decision theory, artificial intelligence, cognitive social theory and behavioral science. As a relatively new area there is a lot of confusion when comparing research efforts from different point of views [35] and therefore there is a need for surveys that record and aggregate efforts done by independent researchers, provide definitions and explain structures and taxonomies of the field from various points of view.

The novelty of this review is that it focuses on web mining in societal benefit areas. We aim to provide the latest developments on applications of web mining, focusing to the need for web mining applications in societal beneficial areas. A number of surveys and textbooks on the field exist in the literature related to web mining [6, 13, 14, 20, 23, 26, 47, 57]. Baldi et. al [6] cover research and theory on aspects of internet and web modeling at the information level based on mathematical, probabilistic, and graphical treatment. Chakrabarti [13] focuses on studies that connect users to the information they seek from the web providing lots of programs. Chen and Chau [14] provide an extended review of how machine learning techniques for traditional information retrieval systems have been improved and adapted for web mining applications. Pal, et. al present an overview of machine learning techniques with focus on web content mining. Studies in [19,23,26,47,57] provide theory and extensive knowledge on web mining categories.

This work is differentiated from the above related work as we will particularly focus on web mining that may help and benefit societal areas in ways of extracting new knowledge, providing support for decision making and empowering valuable management of societal issues. We will also provide

The 9th IEEE International Conference on E-Commerce
Technology and The 4th IEEE International Conference
on Enterprise Computing, E-Commerce and E-Services(CEC-EEE 2007)
0-7695-2913-5/07 $25.00 © 2007 IEEE

for the applications that we present the machine learning techniques that have been used.

The rest of this work is organized as follows. Section 2 describes the taxonomy of web mining focusing on interesting research fields from the users point of view. Section 3 is the main section that presents web mining research on areas of social benefit. Finally in Section 4 we conclude our work.

2. Web Mining Taxonomy

Web mining is divided into three mining categories according to the different sources of data analyzed.

a) Web content mining focus on the discovery of knowledge from the content of web pages and therefore the target data consist of multivariate type of data contained in a web page as text, images, multimedia etc

b) Web usage mining focus on the discovery of knowledge from user navigation data when visiting a website. The target data are requests from users recorded in special files stored in the website's servers called log files.

c) Web structure mining deals with the connectivity of websites and the extraction of knowledge from hyperlinks of the web.

The above taxonomy is now broadly used in web mining and has the origins from Coley et. al [18] who introduced web content mining and web usage mining and Kosala and Blockeel [35], who added web structure mining. Interesting, from the users point of view, research fields in the three categories are described in the next sections.

2.1. Web Content Mining

A well-known problem, related to web content mining, is experienced by any web user trying to find all and only web pages that interests him from the huge amount of available pages. Current search tools suffer from low precision due to irrelevant results [12]. Search engines aren't able to index all pages resulting in imprecise and incomplete searches due to information overload. The overload problem is very difficult to cope as information on the web is immensely and grows dynamically raising scalability issues. Moreover, myriad of text and multimedia data are available on the web prompting the need for intelligence agents for automatic mining.

Agents search the web for relevant information using domain characteristics and user profiles to organize and interpret the discovery information. Agents may be used for intelligent search, for classification of web pages, and for personalized search by learning user preferences and discovering web sources meeting these preferences.

Web content mining is more than selecting relevant documents on the web. Web content mining is related to information extraction and knowledge discovery from analyzing a collection of web documents. Related to web content mining is the effort for organizing the semi-structured web data into structured collection of resources leading to more efficient querying mechanisms and more efficient information collection or extraction. This effort is the main characteristic of the "Semantic Web" [8], which is considered as the next web generation. Semantic Web is based on "ontologies", which are meta-data related to the web page content that make the site meaningful to search engines. Sebastiani study [54] may be used as a source for web content mining.

2.2. Web Usage Mining

Web usage mining research focuses on finding patterns of navigational behavior from users visiting a website. These patterns of navigational behavior can be valuable when searching answers to questions like: How efficient is our website in delivering information? How the users perceive the structure of the website? Can we predict user's next visit? Can we make our site meeting user needs? Can we increase user satisfaction? Can we targeting specific groups of users and make web content personalized to them? Answer to these questions may come from the analysis of the data from log files stored in web servers.

Web usage mining has become a necessity task in order to provide web administrators with meaningful information about users and usage patterns for improving quality of web information and service performance [22, 56, 58]. Successful websites may be those that are customized to meet user preferences both in the presentation of information and in relevance of the content that best fits the user.

Web site personalization is the process of customizing the content and structure of a web site to the specific needs of each user taking advantage of user's navigational behavior [22]. Recommendation systems support web site personalization by tracking user's behavior and recommending similar items to those liked in the past (Content-based filtering), or by inviting users to rate objects and state their preferences and interests so that recommendations can be offered to them based on other users rates with similar preferences (Collaborative filtering), or by asking questions to the user and providing tailored to his needs services according to his answers (Rule-based filtering). Content-based filtering is the most common

The 9th IEEE International Conference on E-Commerce
Technology and The 4th IEEE International Conference
on Enterprise Computing, E-Commerce and E-Services(CEC-EEE 2007)
0-7695-2913-5/07 $25.00 © 2007 IEEE

method for web personalization from server log files and has attracted considerable attention from researchers [42, 43, 46, 55, 57, 61] for constructing user models that represent the behavior of users.

Extracted usage and user behavior patterns may be used in targeting specific groups of users; in various recommendation systems; and in evaluation and reconstruction of the web site to meet design efficiency issues and user satisfaction requirements. Detailed surveys on web usage mining are presented in [23,57].

2.3. Web Structure Mining

Web structure mining is closely related to analyzing hyperlinks and link structure on the web for information retrieval and knowledge discovery. Web structure mining can be used by search engines to rank the relevancy between websites classifying them according to their similarity and relationship between them [35]. Google search engine, for instance, is based on PageRank algorithm [10], which states that the relevance of a page increases with the number of hyperlinks to it from other pages, and in particular of other relevant pages.

Personalization and recommendation systems based on hyperlinks are also studied in web structure mining. Web structure mining is used for identifying “authorities”, which are web pages that are pointed to by a large set of other web pages [20] that make them candidates of good sources of information.

Web structure mining is also used for discovering community networks by extracting knowledge from similarity links. The term is closely related to “link analysis” research, which has been developed in various fields over the last decade such as computer science and mathematics for graph-theory, and social and communication sciences for social network analysis [25, 48, 59]. The method is based on building a graph out of a set of related data [5] and to apply social network theory [59] to discover similarities. Recently Getoor and Diehl [26] introduce the term “link mining” to put special emphasis on the links as the main data for analysis and provide an extended survey on the work that is related to link mining.

Community mining is a major research area on social network analysis that emphasizes on discovering groups of individuals, who by sharing the same properties form a specific community on the web. Domain applications related to web structure mining of social interest are: criminal investigations and security on the web, digital libraries where authoring, citations and cross-references form the community of academics and their publications etc. Detailed survey on web structure mining can be found in [20,26].

3. Web Mining in Societal Benefit Areas

Web mining may benefit those organizations that want to utilize the web as a knowledge base for supporting decision-making. Pattern discovery, analysis and interpretation of mined patterns may lead to better decisions for the organization and for the provided services. E-commerce and E-business were two fields empowered by web mining having lots of applications for increasing sales and doing intelligence business. Lots of web mining applications found in the literature describe the effectiveness of the application from the web administration point of view. The target in these applications is taking advantage of the mined knowledge from the users to increase the benefits for the organization. In our approach we focus on social beneficial areas from web mining, so our point of view is on web mining applications that can help users or group of users. An obvious societal benefit is that web mining research efforts lead to user (or group of users) satisfaction by providing accurate and relevant information retrieval; by providing customized information; by learning about user’s demands so that services can target specific groups or even individual users; and by providing personalized services.

Framing web mining in societal beneficial areas, we need to define what areas or domains are considered as social beneficial, a difficult question which routes to philosophical and quality of life aspects [19]. The possible number of domains is large. Material well being, health, productivity, intimacy, safety, community and emotional well being are domains suggested in [19] that are related to quality of life. In [37] the domains: individual and families, community participation, environment, health, education, physical security, economic security, culture and leisure, social services, and government are indicators investigated for quality of life.

In this work we frame social beneficial areas based on meta-research results where we first identified web mining research in various areas that may be of social interest. Then we tried to group the identified research into related categories. The result for our societal beneficial areas investigation divides web mining research into six social beneficial areas: a) E-learning, where educational improvement is the benefit; b) E-Government, where improvement of government services to citizens is the benefit; c) E-politics and e-democracy, where community participation is the benefit d) Helpdesks and Recommendation systems, where improvement of social services is the benefit; e) Digital Libraries, where improvement of productivity by diffusion of ideas is the benefit; f) Security and

The 9th IEEE International Conference on E-Commerce
Technology and The 4th IEEE International Conference
on Enterprise Computing, E-Commerce and E-Services(CEC-EEE 2007)
0-7695-2913-5/07 $25.00 © 2007 IEEE

crime investigation, where public safety is the benefit. The first three areas are considered to be part of what is widely known as e-services. We will see in the next sessions how current web mining research is related to the above social beneficial areas.

3.1. E-Learning

Web mining can be used for enhancing the learning process in e-learning environments. Applications of web mining to e-learning are usually web-usage based applications. Bellaachia et. al. [7], introduce a framework, where they use logs to analyze the navigational behavior and the performance of e-learners so that to personalize the learning content of an adaptive learning environment in order to make the learner reach his learning objective. Zaiane [62] studies the use of machine learning techniques and web usage mining to enhance web-based learning environments for the educator to better evaluate the learning process and for the learners to help them in their learning task. Students' web logs are investigated and analysed in Cohen and Nachmias [17] in a web-supported instruction model.

3.2. E-Government

The process by organizations that interact with citizens for satisfying user (or group of users) preferences leads to better social services. The major characteristics of e-government systems are related to the use of technology to deliver services electronically focusing on citizens needs by providing adequate information and enhanced services to citizens to support political conduct of government. Empowered by web mining methods e-government systems may provide customized services to citizens resulting to user satisfaction, quality of services, support in citizens decision making, and finally leads to social benefits. Such social benefits much rely on the organization's willingness, knowledge and ability to move on the level of using web mining. The e-government dimension of an institution is usually implemented gradually. E-government maturity models [33, 36] describe the online stages an organization goes through time, becoming more mature in using the web for providing better services to citizens. The maturity stages start from the organization's first attempt to be online aiming at publishing useful to citizens information and move to higher maturity stages of being interactive, making transactions and finally transforming the functionality of the organization to operate their business and services electronically through the web. Maturity stages described in literature doesn't have a web mining dimension, which we consider that should be the climax in maturity stages. Riedl [51] states that by using interviews and web mining the actual access to information by citizens should be tracked, analyzed and used for the redesign of e-government information services. E-Government literature reveals that only recently web mining has attracted researchers in e-government applications. Fang and Sheng [24] present a web mining approach for designing better web portal for e-government. Hong and Lee [32] propose an intelligent web information system of government based on web usage mining to help disadvantaged users make good decisions-making for their profit improvements. In the health sector improvements of health services by quality labelling of medical web content are investigated in [40].

3.3. E-Politics and E-Democracy

E-politics provides political information and politics "on demand" to the citizens improving political transparency and democracy, benefiting parties, candidates, citizens and the society. Election campaigners, parties, members of parliament and members of local governments on the web are part of e-politics. Despite the importance of e-politics in democracy there is limited web mining methods to meet citizen needs. We identified in the literature research that only refers in mining political communities, being web-structure based efforts. Link analysis has been used to estimate the size of political web graphs [2], to map political parties network on the web [1] and to investigate the U.S. political Blogosphere [3]. Political web linking is also studied by Foot et. al. [25] during the U.S. Congressional election campaign season on the web.

3.4. Helpdesks and Recommendation Systems

Recommendation systems are based on user modeling mainly derived from content-based learning or from collaborative filtering [63]. Content–based learning uses a user's past usage behavior and acts as an indicator of his future behavior. Collaborative filtering is based on ratings of user favors, like rating music or movies, so that rating history of a user can be associated with similar preferences of other users. So a user is classified in a user model, where recommendations can be addressed to him according to favors of other people from the specific classified user model. Hybrid recommendation systems that take benefits from both collaborative filtering and content-based learning have been also investigated in literature

The 9th IEEE International Conference on E-Commerce
Technology and The 4th IEEE International Conference
on Enterprise Computing, E-Commerce and E-Services(CEC-EEE 2007)
0-7695-2913-5/07 $25.00 © 2007 IEEE

[39, 41, 52]. Martin-Guerrero et. al. [39] propose a recommendation model for predicting user preferences based on common clustering algorithms in a citizen web portal. Clustering and collaboration filtering is used in Hayes et. al. [30] for a blog recommendation system. A blog is a journal-style website usually written by a single user, where entries are presented in a reverse chronological order. ReferralWeb [34] is a project that mines social networks from the web by using collaborative filtering for identifying experts that could answer questions asked by individuals. Nasrui and Pavuluri [45] using neural networks provide accurate web recommendations based on a committee of predictors. Yao et. al. [64] created PagePrompter an agent-based recommendation model that helps users navigate a website by analysing and learning from web usage mining and user behaviour. The interesting part of Pageprompter is that it combines three different machine learning techniques: association rules, clustering and decision trees for achieving its task. Pierrakos et. al [49] use clustering applied to web usage mining for creating community specific directories to offer users a more personalized view of the web according to their preferences and may be assisted by using these directories as starting points on their navigation. Scheffer [53] created an email-answering assistant by using semi-supervised text classification

3.5. Digital Libraries

Digital libraries provide precious information distributed all around the world without necessarily having the need to be physically present in a traditional library building. Web mining research aiming to better services on digital libraries have been identified in literature. Adafre and Rijke [4] use clustering for discovering missing hypertext links in Wikipedia, the largest encyclopedia on the web that is created and modified by many volunteer authors. Web mining on wikipedia is also investigated by Gleim, et. al. [28]. Bhattacharya and Getoor [9] use clustering for detecting group of entities, like authors, from links and resolving the co-reference problem of multiple references to the same paper in autonomous citation indexing engines, like CiteSeer [27]. CiteSeer is an important resource for computer scientists for searching electronic versions of papers. Cai et. al [11] work also is based on another well-known digital library in computer science community, the DBLP library at http://dblp.uni-trier.de. Their work is related to machine learning feature extraction algorithms in order to discover hidden communities in heterogeneous social networks. Graph analysis for discovering web communities can be modeled by using Bayesian networks as demonstrated by Goldenberg and Moore [29] for identifying co-authorship networks. A content-based book recommendation system is proposed by Money, and Roy [44] based on web pages of the Amazon on-line digital store. Large portals with news updated frequently per day consist of rich information and may be considered as part of digital libraries in the way that newspaper articles are indexed and available to readers in traditional libraries. News sites are large portal sites that increase their content on a daily basis. For such sites, the interpretation of web content to meaningful content that can be classified into semantic categories in order to make both information retrieval and presentation easier for individuals and group of users is very important [22]. Liu et. al [38] use fuzzy clustering to identify meaningful news patterns from web news stream data.

3.6. Security and Crime Investigation

Web mining techniques are used for identifying cyber-crime actions like internet fraud and fraudulent websites, illegal online gambling, hacking, virus spreading, child pornography distribution and cyber-terrorism. Chen, Chung, et. al., [15] note that clustering and classification techniques can reveal identities of cyber-criminals, whereas neural networks, decision tress, genetic algorithms and support vector machines can be used to crime patterns and network visualization. Chen, Qin, et. al. [16] provide a detailed study on methods against terrorist groups on the web for predictive modeling, terrorist network analysis, visualization of terrorists' activities, linkages and relationships. Wu et. al. [60] based on user's online activities use principal cluster analysis to identify a small number of major topics from millions of navigational data in an approach that can be useful in security against terrorism. Do et. al. [21] implemented a system that can be beneficial to safe internet browsing in school, home and workplace. The system monitors and filters web access by applying web mining for performing web data classification in order to classify web data in a "whitelist" of allowed pages or blacklist of blocked web pages. Social Networks extracted from instant messaging by using clustering is investigated by Resig, et. al. [50].

4. Conclusions

We provided a survey on web mining focusing on current web mining research in societal benefit areas identifying that most of this research has been recently developed. Therefore, one of the current trends of web mining is toward the connection between intelligent web services and applications of social benefits, which brings to work closer scientists from various disciplines.

The E-Services areas of societal interest that have been included in our work are: E-learning, E-Government, E-Politics and E-Democracy. Only recently web mining applications have been related to these three fields. Current web mining research on E-learning is based on web usage mining as the focus has been on how the student performs. Current web mining research focus on E-Politics is based on web structure mining to identify political groups. It seems that the fields of E-services and web mining have recently met each other. Although the fields have not been yet empowered by web mining the challenge is open to seen more research on these fields leading to more benefits to society.

Apart of E-services, we surveyed web mining research on fields that are of societal interest due to the benefits a user or a group of users may gain. Such systems are Helpdesks and Recommendation systems; Digital Libraries; and Security and Crime Investigations systems. These fields have been under the focus from more researchers and for longer time than the E-Services fields. Thus, these fields have been empowered more by the web mining research than the E-Services fields.

Societal benefits areas are expected to be further empowered in the future by the combination of web mining and semantic web [8], where websites have meaningful content to search engines and to machine learning systems.

References

[1] Ackland, R., and Gibson, R. (2004). Mapping Political Party Networks on the WWW. *Australian Electronic Governance Conference*, Melbourne.

[2] Ackland, R. (2005). Estimating the Size of Political Web Graphs, revised paper presented to ISA Research Committee on Logic and Methodology Conference, Retrieved from http://acsr.anu.edu.au/staff/ackland/papers/political_web_graphs.pdf (15th, January, 2007)

[3] Ackland, R. (2005). Mapping the U.S. Political Blogosphere: Are Conservative Bloggers More Prominent?, *paper presented to BlogTalk Downunder* , Sydney, Retrieved from http://acsr.anu.edu.au/staff/ackland/papers/polblogs.pdf (15th, January, 2007).

[4] Adafre, S.F., and Rijke, M.d. (2005). Discovering Missing Links in Wikipedia. *Proceedings of the 3rd International Workshop on Link Discovery*, ACM Press, 90-97.

[5] Badia, A., and Kantardzik, M. (2005). Graph Building as a Mining Activity: Finding Links in the Small. *Proceedings of the 3rd International Workshop on Link Discovery*, ACM Press, 17-24.

[6] Baldi, P., Frasconi, P., and Smyth, P. (2003). *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, Wiley, West Sussex, UK.

[7] Bellaachia, A., Vommina, E., and Berrada, B. (2006). Minel: A Framework for Mining E-learning Logs, *Proceedings of the 5th IASTED International Conference on Web-based Education*, Puerto Vallarta, Mexico, 259-263.

[8] Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5):34-43.

[9] Bhattacharya, I., and L.Getoor (2004). Deduplication and group detection using links, in *SIGKDD Workshop on LinkAnalysis and Group Detection*, Seattle, WA, USA.

[10] Brin, S., and Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Proceedings of the 7th International World Wide Web Conference*, Elsevier Science, New York, 107-117.

[11] Cai, D., Shao, Z., He, X., Yan, X., and Han, J. (2005). Minning Hidden Community in Heterogeneous Social Networks. *Proceedings of the 3rd International Workshop on Link Discovery*, ACM Press, 58-65.

[12] Chakrabarti, S., (2000). Data Mining for Hypertext: A Tutorial Survey. *ACM SIGDDD Explorations*, 1(2):1-11.

[13] Chakrabarti, S., (2003). *Mining the Web: Discovering Knowledge from Hypertext Data*, Morgan Kaufmann Publishers, San Francisco, USA.

[14] Chen, H., and Chau, M. (2004). Web Mining: Machine Learning for Web Applications, *Annual Review of Information Science and Technology (ARIST)*, 38:289-329.

[15] Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., and Chau, M. (2004). Crime Data Mining: A General Framework and Some Examples. Computer, 37(4):50-56.

[16] Chen, H., Qin, J., Reid, E., Chung, W., Zhou, Y., Xi, W., Lai, G., Bonillas, A. and Sageman, M., "The Dark Web Portal: Collecting and Analyzing the Presence of Domestic and International Terrorist Groups on the Web, *Proceedings of the 7th International Conference on Intelligent Transportation Systems (ITSC),* Washington D.C., October 3-6, 2004

[17] Cohen, A., and Nachmias, R. (2006). A Quantitative Cost Effectiveness Model for Web-Supported Academic Instruction. *The Internet and Higher Education*, 9(2):81-90.

[18] Cooley, R., Mobasher, B. and Srivastava, J., (1997). Web Mining: Information and Pattern Discovery on the World Wide Web, *9th International Conference on Tools with Artificial Intelligence(ICTAI '97)*, New Port Beach, CA, USA, IEEE Computer Society, 558-567.

[19] Cummins, R.A. (1996). The Domains of Life Satisfaction: An Attempt to Order Chaos, Social Indicators Research, 38:303-328.

[20] Desikan, P., Srivastava, J., Kumar, V., and Tan, P.N. (2002). *Hyperlink Analysis: Techniques and Applications*, Technical Report (TR 2002-0152), Army High Performance Computing Center.

[21] Do, T.D., Chang, K., and Hui, S.C. (2004). Web Mining for Cyber Monitoring and filtering, in *Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems*, Singapore, Vol. 1:399-404.

[22] Eirinaki, M., and Vazirgiannis, M. (2003). Web Mining for Web Personalization. *ACM Transactions on Internet Technology*, 3(1):1-27.

[23] Faca, F.M., and Lanzi, P.L. (2005). Mining Interesting Knowledge from Weblogs: A Survey, *Data Knowledge Engineering*, 53(3):225-241.

[24] Fang, X., Sheng, O.R.L. (2005). Designing a Better Web Portal for Digital Government: A Web-Mining Based Approach. In Proceedings of the 2005 National Conference on Digital Government Research, Atlanta, 277-278.

[25] Foot, K., Schneider, S., Dougherty, M., Xenos, M., and Larsen, E. (2003). Analyzing Linking Practices: Candidate Sites in the 2002 U.S. Electoral Web Sphere. *Journal of Mediated Communication*, 8(4).

[26] Getoor, L. and Diehl, C.P. (2005). Link Mining: A Survey. *ACM SIGKDD Explorations Newsletter*, 7(2):3-12.

[27] Giles, C.L., Bollacker, K., and Lawrence, S. (1998). CiteSeer: An Automatic Citation Indexing System. *Proceedings of the 3rd ACM Conference on Digital Libraries*, 89-98.

[28] Gleim R., Mehler A., Dehmer M. (2006). Web Corpus Mining by Instance of Wikipedia, Proceedings *of the EACL 2006 Workshop on Web as Corpus*, Trento, Italy.

[29] Goldenberg, A., and Moore, A.W. (2005). Bayes Net Graphs to Understand Co-authorship Networks? *Proceedings of the 3rd International Workshop on Link Discovery*, ACM Press, 1-8.

[30] Hayes, C., Avesani, P., and Veeramachaneni, S. (2006). An Analysis of Bloggers and Topics for a Blog Recommender System, *in proceedings of the Workshop on Web Mining , 7th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD)*, Berlin, Germany.

[31] Herrera-Viedma, E., and Pasi, G. (2006). Soft Approaches to Information Retrieval and Information Access on the Web: An Introduction to the Special Topic Section, *Journal of the American Society for Information Science and Technology*, 57(4):511-514.

[32] Hong, G.H., and Lee, J.H. (2005). Designing an Intelligent Web Information System of Government Based on Web Mining, *Lecture Notes in Computer Science*, 3614:1071-1078.

[33] Irani, Z., Al-Sebie, M., Elliman, T. (2006). Transaction Stage of e-Government Systems : Identification of its Location & Importance, *in Proceedings of the 39th Hawaii International Conference on System Sciences,* Hawai, USA.

[34] Kautz, H., Selman, B., and Shah, M. (1997). Referral Web: Combining Social Networks and Collaborating Filtering. *Communications of the ACM*, 40(3):63-65.

[35] Kosala, R., and Blockeel, H., (2000). Web Mining Research: A Survey, *ACM* 2(1):1-15.

[36] Lappas, G., and Yannas, P. (2006). A Framework to Evaluate Political Party Websites, in *Proceedings of the 4th International Conference on Politics and Information Systems: Technologies and Applications*, Orlando, Florida, USA, Vol. II: 226-231.

[37] Legowski, B. (2000). *A Sampling of Community and Citizen Driven Quality of Life/Societal Indicator Projects*, Canadian Policy Research Networks Report.

[38] Liu, J.W., Yu, S.J., and Le, J.J. (2005). Online Mining Dynamic Web News Patterns Using Machine Learn Methods, *Lecture Notes on Artificial Intelligence*, 3614:462-465.

[39] Martin-Guerrero, J.D., Palomares, A., Balaguer-Ballester, E., Soria-Olivas, E., Gomez-Sanchis, J., and Soriano-Asensi, A. (2006). Studying the Feasibility of a Recommender in a Citizen Web portal Based on User Modeling and Clustering Algorithms. *Expert Systems with Applications*, 30:299-312.

[40] Mayer, M.A., Karkaletsis, V., Stamatakis, K., Leis, A., Villarroel, D., Thomeczek, C., Labsky, M., Lopez-Ostenero, F., and Honkela, T. (2006). MedIQ-Quality Labelling of Medical Web Content Using Multilingual Information Extraction, *Studies in Health Technology and Informatics*, 2006, 121: 183-190.

[41] Melville, P., Mooney, R.J., and Nagarajan, R. (2002). Content-boosted Collaborative Filtering for Improved Recommendations, *Proceedings of the 18th National Conference on Artificial Intelligence*, 187-192.

[42] Mobasher, B., Jain, N., Han, E., and Srivastava, J. (1996). *Web Minning: Pattern Discovery from WWW Transaction*. Technical Report TR-96050, Department of Computer Science, University of Minnesota, Minneapolis, Retrieved from http://citeseer.ist.psu.edu/mobasher96web.html (5th, January, 2007).

[43] Mobasher, B., Cooley, R., and Srivastava, J. (1999). Creating Adaptive Web Sites Through Usage Based Clustering of URLs. *In IEEE Knowledge and Data Engineering Exchange Workshop* (KDEX99), Chicago, Illinois, USA.

[44] Mooney, R.J., and Roy, L. (2000). Content-based Book Recommending Using Learning for Text Categorization. *Proceedings of the 5th ACM Conference on Digital Libraries*, ACM Press, 195-204.

[45] Nasraoui, O., and Pavuluri, M. (2004). Complete this Puzzle : A Connectionist Approach to Accurate Web Recommendations Based on a Committee of Predictors, in *Proceedings of the 6th WEBKDD Workshop*, Seattle, USA.

[46] Ngu, D.S.W., and Wu, X. (1997). Sitehelper: A Localized Agent that Helps Incremental Exploration of the World Wide Web. Computer Networks, 29(8-13):1249-1255.

[47] Pal, S., Talwar, V., and Mitra, P. (2002). Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions, IEEE Transactions on Neural Networks, 13(5):1163-1177.

[48] Park, H.W. (2003). Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web, *Connections*, 25(1), 49-61.

[49] Pierrakos, D., Paliouras, G., Papatheodorou, C., Karkaletsis, V., and Dikaiakos, M. (2003). Web Community Directories : A New Approach to Web Personalization, *Lecture Notes on Artificial Intelligence*, 3209 :113-129

[50] Resig, J., Dawara, S., Homan, C.M., and Teredesai, A. (2004). Extracting Social Networks from Instant Messaging Populations, *LinkKDD'04*, Seattle, Washington, USA.

[51] Riedl, R. (2003). Design Principles for E-Government Services. *In Proc. eGov Day 2003,* Vienna, Austria.

[52] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2000). Analysis of Recommendation Algorithms for E-Commerce, *Proceedings of the ACM Conference on Electronic Commerce*, 158-162.

[53] Scheffer, T. (2004). Email Answering Assistance by Semi-supervised Text Classification. *Intelligent Data Analysis*, 8(5), 2004.

[54] Sebastiani. F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1):1–47.

[55] Spiliopoulou, M., Pohle, C., and Faulstich, L. (1999). Improving the Effectiveness of a Web Site with Web Usage Mining, *In Proceedings of WEBKDD99*. San Diego, CA, USA, 142-162.

[56] Spiliopoulou, M., and Pohle, C. (2001). Data Mining for Measuring and Improving the Success of Web Sites. *Data Mining and Knowledge Discover*, 5(1-2):85-114.

[57] Srivastava, J., Cooley, R., Deshpande, M., and Tan, P. (2000). Web Usage Mining:Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, 1:12-23.

[58] Wang, X., Abraham, A., and Smith, K. (2005). Intelligent Web Traffic Mining and Analysis. *Journal of Network and Computer Applications,* 28:147-165.

[59] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press.

[60] Wu, H., Gordon, M., DeMaagd, K., and Fan, W. (2006). Mining Web Navigations for Intelligence. *Decision Support Systems*, 41:574-591

[61] Wolfgang, G., and Lars, S. (2000). Mining Web Navigation Path Fragments. In *Workshop on Web Mining for E-Commerce (KDD2000),* Boston, MA, USA, 105-110.

[62] Zaiane, O.R. (2001). Web Usage Mining for a Better Web-Based Learning Environment. *in Proceedings of Conference on Advanced Technology for Education*, pp, Banff, Alberta, Canada, 60-64.

[63] Zukerman, I., and Albrecht, D. (2001). Predictive Statistical Models for User Modeling. *User Modelling and User Adapted Interaction*, 11:5-18.

[64] Yao, Y.Y., Hamilton, H.J., and Wang, X. (2002). PagePrompter: An Intelligent Agent for Web Navigation Created Using Data Mining Techniques, *Lecture Notes In Computer Science,* 2475: 506-513