Available online at www.elixirpublishers.com (Elixir International Journal)

Advanced Engineering Informatics

Elixir Adv. Engg. Info. 37 (2011) 3948-3952

ISSN: 2229-712X

Virtual information environments

Sundar.C[1], Mayilvelkumar.P[2] and Dennis Ebenezer.D[1]
[1]Christian college of Engineering and Technology, Oddanchatram
[2]Karpagam Institute of Technology, Coimbatore.

ARTICLE INFO

Article history:
Received: 17 May 2011;
Received in revised form:
20 July 2011;
Accepted: 30 July 2011;

Keywords
Alert system,
Environmental scanning,
Hyperbolic self-organizing maps,
Information foraging.

ABSTRACT
The recovery of significance topics and associated sources from environment where the information is readily available. In this project, fusion system has been developed to handle the information assembly method. Hierarchal Hyperbolic SOM (H_2SOM) network is used in this system to configure the information background for information seeking process and to recover the information for the significant topic. Activation network has been used to choose the related information with reference to the existing knowledge. This network is used to choose the appropriate information from the source. This project use association rule of Data mining to examine the recognized topics. Association rule will perform numerical testing to examine the recognized topics. This system will be implemented in several industries and most applicable domain is hospital industry. Information overload has been monitored by this system. In this system information assembly method will be achieved effectively.

Introduction

Scanning effectively and identifying key information quickly are the two most crucial skills for business managers who want to succeed in today's information intensive environments. Simon [1, pp. 40-41] highlights this problem in the following way: "What information consumes is rather obvious: it consumes the attention of its recipients. Hence, a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it." Consequently, managers need ways to understand their business environment as well as to integrate this understanding into their planning and decision-making processes. The term Environmental Scanning (ES) refers to methods to achieve this integrative understanding. In our exemplarily considered, application domain of the hospitality industry, as well as in almost all other industries, ES activities are hampered by an information overload caused by the dramatic increase of relevant documents and messages emitted by an increasing number of information sources [2], [3]. The need for spending efforts, time, and resources in ES activities arises from the competition in recognizing new developments before a competitor does in order to set up superior business strategies and tactics already adjusted to the new developments.

One of the information environments becoming increasingly important with respect to ES is the World Wide Web (WWW). While information quality remains a major challenge, the sheer plethora of information sources plus the often available time stamp indicating when a piece of information became available makes the WWW an ideal environment to pick up crucial changes in the business sector. The information seeking process is heavily affected by the mental model of the manager. From a cognitive perspective, in a rough outline, the process of ES can be split into three stages [4] as follows:

i. The discovery of new information should ideally be largely unguided by the a priori knowledge and predisposition of the manager; typically conceptualized in mental models, in order to prevent any restriction to already known areas and phenomena in the business environment.

ii. During the expansion of already existing knowledge of the business world, the manager follows his or her well-defined information interests and thus, focuses more on particular aspects that are considered to be worth investigating in more detail.

iii. In the third stage, there is a shift from detection to monitoring of developments in the business environment. Consequently, the mental model is updated to changes in already explored areas of the environment.

A successful support of the above ES process must account for the limited information processing capacity of humans: The first stage is the crucial stage triggering any subsidiary ES activities. Consequently, not the ability to scan but the ability to notice is the first step in augmenting the mental model of the manager to the business environment [5]. Against this background, the main objective of automated ES systems is to facilitate sensitive and context dependent reductions of the information overload that makes up the environment in which modern businesses must operate [3]. It should be noted that these reductions have to follow the idiosyncratic state of mind and information needs of the manager in order to grasp the relevant information of the business environment. So far, to the best of our knowledge, existing automatic ES systems lack in at least one of the three above modules: Wei and Lee [2] present a promising system that facilitates the detection of new events in document streams in the context of ES. However, with respect to the above ES process, the system does not support the discovery stage, since a pre specified set of event topics and their respective properties have to be defined beforehand by the manager. Moreover, the system does not present the results in any visual context. Thus, the detection of weak signals spread over various documents is not facilitated. Other ES systems concentrate rather on processes of information gathering and displaying (cf. [6]) but fall short of applying sophisticated methods of information reduction that exceed standard query-

Tele:
E-mail addresses: mayilvelkumar@gmail.com, sundarc007@yahoo.com

based filtering. For these reasons, the discovery of new phenomena is hardly supported in existing ES system approaches [4]. Moreover, no system supports the entire ES processor, therefore, assists the complete progression from searching and detecting new phenomena to regularly observing these on a daily basis.

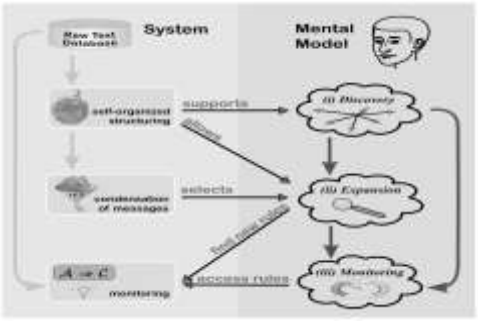

Fig 1 ES Process

To that effect, this paper outlines an integrated system that contributes to effective ES, in terms of learning about the most relevant developments in the business environment with minimal efforts; by means of a text mining architecture utilizing three modules that map all three cognitive stages of the ES process (sees Fig. 1).

1. Discovery: In a first step, raw textual data from an information source are organized by a self-organizing clustering process. As an innovative clustering methodology, we introduce a hierarchically growing neural network "H2SOM", which is based on the self organizing map (SOM) principle. A focus and context- aware graphical user interface, which displays the cluster structure in a highly condensed interactive visualization scheme, allows for the detection of cluster structures with respect to semantic contents and temporal correlations.
2. Expansion: Once the decision-maker has identified potentially interesting topics, there might still be hundreds of single messages related to them. By defining an information infrastructure describing the manager's information needs, the Information Foraging Theory (IFT) allows further condensation to just a few essential messages, which might then be easily "digested" by the manager.
3. Monitoring: Knowledge generated in phases 1 and 2 is used to formulate association rules consisting of relevant terms in both the rules' antecedents and consequents. We outline how statistical testing of these rules gives insights into the development of topics in the course of time and how they might be utilized to implement an automatic alert system.

Managerial Information Seeking

Situation Awareness

Regardless of any context, deciding involves judging by means of assessing information and appraising decision alternatives with respect to consequences for all possible worlds and future contingencies [7], but in changing environments, this needs the continuous updating of information on the business environment. According to the Sacagawea principle, computational tools should support active organization of information, active search of information, active exploration of information reflection of meanings as well as the evaluation and choice among sequences of alternatives [8]. The core of this principle is situation awareness by means of updating the mental model of the ongoing situation. So far, the managerial information needs are quite similar to the information needs of pilots, physicians, or weather forecasters. They all have to deal with complex information environments, where complexity arises from a high number of variables, their independencies, and the emergence of meanings, which is not obvious by considering information fragments [9]. Successful tools have been designed to cope with different kinds of "well structured problems," but managers have been found to be wary of using analytical tools [10] in more complex decision situations. Of course, they have a good reason for their skepticism: Superior management needs creativity by means of forsaking the conventional views and being open-minded to different assessments of information. Thus, managers need the opportunity to explore information interactively and to combine previously unrelated facts (e.g., with respect to their originator or the date of origin). Clearly, being open-minded does not imply being mindless. Once a manager identifies a topic relevant to his or her business decisions, he or she is interested in precise information and, particularly, in changes of the relations of facts. In contrast to pilots or weather forecasters, managers get a suitable share of their information by qualitative textual information, e.g., announcements of competitors or news services, rather than quantitative measurements such as temperature and air pressure. To make the situation even worse, the antecedents of this information are unclear and hardly described by simple rules. Even in a retrospective consideration, the causality of market phenomena is often subject to controversial debates. Consequently, ascertaining the relevance of individual facts is difficult.

Application Domain

To demonstrate our methodology, we draw on the example of 2,314 documents obtained from the Internet-based hospitality industry newsletter, ehotelier.com. In this industry, the managers have to cope with all the challenges outlined above. Their performance is highly sensitive to events originating from outside the industry, like important sport events, exhibitions, and fairs, but also catastrophes and political and economic changes. Thus, several information services aim to provide the managers with all kinds of relevant information.

Following standard procedure in information retrieval, we build a vector-space model; commonly referred to as the bag of words. We extracted a dictionary with 7,457 entries after stop word removal; applying Porter's stemming and feature selection by taking only those words into account that appear in more than three messages. This dictionary was used to construct feature vectors, which are weighted with the classic term frequency inverse document frequency scheme.

Methods

Hierarchically Growing Hyperbolic Self-Organizing Maps:

Hyperbolic Multi-Dimensional Scaling (HMDS) and Hyperbolic Self-Organizing Maps (HSOM) employ the extraordinary advantages of the hyperbolic plane (H2). The underlying space grows exponentially with its radius around each point - ideal for embedding high-dimensional data, the Poincare model of the IH2 exhibits a fish-eye perspective with a focus area and a context preserving surrounding, the mouse binding of focus-transfer allows intuitive interactive navigation. The HMDS approach extends multi-dimensional scaling and generates a spatial embedding of the data representing their dissimilarity structure as faithfully as possible. It is very suitable for interactive browsing of data object collections, but calls for batch precomputation for larger collection sizes. The HSOM is an extension of Kohonen Self-Organizing Map and generates a partitioning of the data collection assigned to an IH2 grid. The

algorithm complexity is linear in the collection size; the data browsing is rigidly bound to the underlying grid.

Information Foraging Theory

Information Foraging Theory is an approach to understand how strategies and technologies for information seeking, gathering, and consumption are adapted to the flux of information in the environment. The theory assumes that people, when possible, will modify their strategies or the structure of the environment to maximize their rate of gaining valuable information. Field studies inform the theory by illustrating that people do freely structure their environments and their strategies to yield higher gains in information foraging. The theory is developed by an adaptation (rational) analysis of information foraging problems and a detailed process model (ACT-IF). The adaptation analysis develops information patch models, which deal with time allocation and information filtering and enrichment activities in environments in which information is encountered in clusters (e.g., bibliographic collections),information scent models which address the identification of information value from proximal cues, and information diet models which address decisions about the selection and pursuit of information items. Our adaptive success depends to a large extent on a vast and complex tributary of cultural tasks that engage our physical and social environments. These tasks require information-gathering, sense-making, decision-making, and problem-solving strategies. We are interested in understanding these information-gathering and sense-making strategies from an evolutionary ecological perspective, treating adaptations to the flux of information in the cultural environment in the same manner as biologists study adaptations to the flux of energy in the physical environment. Information Foraging Theory in many ways analogous to evolutionary ecological explanations.

Event detection

Event detection technique that identifies the onset of new events from streams of news stories would facilitate the process of organization's environmental scanning. Traditional feature-based event detection techniques cannot capture the genuine properties of an event contained in a news story and cannot support event categorization and news stories filter. This include the development of an information extraction-based event detection (NEED) technique that combines information extraction and text categorization techniques to address the problems inherent to traditional feature-based event detection techniques. Using a traditional feature-based event detection technique (INCR) as benchmarks, the empirical evaluation results showed that the proposed NEED technique improved the effectiveness of event detection measured by miss and false alarm rates. An event detection technique based on the information extraction approach, called information Extraction-based Event Detection (NEED) technique. The proposed technique will empirically be evaluated, using a traditional event detection technique as benchmarks. Feature-based event detection techniques can be solved by performing event detection based on event property values embedded in news stories rather than features appearing in news stories.

Performance Evaluation

For many scenarios of ES, large amounts of data have to be taken into account. Especially for the scanning and evaluation of Web sources, it is of premier importance that the underlying algorithms scale favorably with the ever growing amount of information. The first component of our system has to deal with the largest amount of data: It has to structure the raw unfiltered stream of information that is directly hitting the system during the ES process. The SOM as introduced by Kohonen more than 20 years ago [17] scales linearly with the problem size and has proven its applicability to real-world data sets, as the WEBSOM project impressively has shown [33]. Though the SOM scales well, the computational effort to train very large maps is still considerable: with linearly increasing map area; which is directly linked to the semantic resolution the SOM is able to achieve the number of nodes in an SOM increases quadratic ally. Therefore, leading to potentially very long training times.

Fast Tree Search Capability of the H2SOM

The peculiar, intrinsically "uniformly hierarchical" structure of the hyperbolic grid that is employed in the first component of our ES system offers an intriguing possibility to significantly accelerate the most time-consuming step in an SOM: We can approximate the global search for the winner unit by a fast tree search, taking as the search root the initial center node of the growth process and then following the "natural" hierarchical structure in the hyperbolic grid. Starting from this node, we recursively determine the k best matching nodes among its nb neighbors until we reach the periphery. For k ¼ 1, this will generate a path with comparisons, instead of a global search in a standard SOM consisting of N nodes.

Assessment of Association Rules in the Course of Time allows for an "SF search." Here, we choose a tree branching factor of $k > 1$ for the first structural level and truncate it to k ¼ 1 for all search steps beyond. This results in an exploration of the search space into different directions but still scales very favorably.

The initial training of the neural network uses an iterative scheme where the search for the best match unit is repeated several thousand times. For the here reported results, we have trained the map with 20N steps with N being the number of documents. Consequently, for the example of the 2,314 documents from the ehotelier newsletter, the hierarchical organization process with the H2SOM takes approximately 7 minutes on a Pentium 4-based PC. A similar sized standard Euclidean SOM would need more than 9 hours for completion on this machine. Note that the initial training is just performed once for each new data source. Once the network has learned its internal representation, it can be used without further computational resources during the discovery phase of our ES system. For changing information sources such as news feeds, the beneficial runtime of several minutes allows for a regular update of the H2SOM, e.g., on a half-daily basis.

Usability Evaluation

In an empirical study, the IFT module outlined in Section 3.2 has proven to outperform human experts in assessing and selecting of relevant documents [3]. For module 3, the empirical application presented in Section 3.3 already provides a face validation of the methodology. A further investigation of the usability of this module would need an external criterion. Noteworthy, the rules have to match the idiosyncratic state of mind and information needs of the human judges. Thus, the result of a usability study would rather reflect the fit of the data under consideration with respect to the manager's needs than the contribution of the methodology in itself.

We believe that one important surplus of our ES system originates from the interactive 3D visualization of the H2SOM, which facilitates a multiperspective approach to the business environment that is fundamental for successful ES. Neugarten

stresses this relation in the following way [5, p. 99]: "Different viewpoints and perspectives should be sought deliberately so as to crosscheck information, just as our two eyes use parallax to allow us to triangulate depth information." Although there has been extensive research on designing and implementing focus and context techniques, few empirical studies have investigated the gains in usability, speed, and performance when using hyperbolic tree browsers. These studies come up with mixed results:

Lamping et al. [34] could not find significant differences between a conventional 2D scrolling tree browser and their hyperbolic tree browser. Czerwinski and Larson [35] compared the latter with Microsoft's Explorer and they, too, could not detect any significant differences in user performance. In contrast, the experiments in [36] and [37] found that subjects using the hyperbolic tree browser gained substantially superior results with respect to user performance. These contradicting results can be at least partly explained by the complexity of the conducted user tasks. In the latter two studies, subjects had to solve cognitively more complex tasks. Obviously, the discovery stage of the ES process, comprising the detection of new information in the business environment from various information sources, imposes an ill-defined problem of high complexity.

Discussion and Conclusions

This paper seizes the challenge of an integrative perception of the information environment by aligning the technical features of an ES system to a cognitive process perspective. Breaking down this process into three steps, we introduce technologies for supporting the ES process in its entirety by means of a hybrid intelligent system. In a first step, the discovery of new information, the H2SOM generates a hierarchically structured information environment suitable for an interactive visual exploration. The annotation of the information patches and the spatial relation to neighboring patches on the "document landscape" provide an intuitively and easily graspable visualization of the underlying structure of the topics as well as the changes in the discussion. It enables different perspectives for discovering and exploring new phenomena in the business environment [5]. Our usability evaluation suggests that particularly the graphical user interface endows even less computer-skilled users with an easy to use operation that motivates them to browse through the information environment. The hierarchically ordered focus and context visualization scheme is not limited to a fixed partition of the information environment but endows the user with an interactive drill down in information patches to be divided into smaller ones. Thus, we argue that our system facilitates both achieving a structured understanding and approaching more information in a given time.

In the second step, the expansion of knowledge, the IFT enables a context-sensitive allocation of the manager's time to the most relevant information sources relating to a topic of interest. Thereby, the manager's information needs and his or her prior knowledge are taken into account by means of the information infrastructure Q. The example of the terror attacks on Bali indicates that by "digesting" the selected (sub) cluster the IFT is able to significantly reduce the amount of information the manager actually has to read without losing important information with respect to Q. Similar results on the test corpus (not reported here due to space limitations) were obtained with information on hotel mergers, the outbreak of the SARS virus, and the war in Iraq. Thus, the IFT module supports the user in quickly accessing the information environment restructured with the H2SOM. Moreover, the user is not in danger of overlooking relevant aspects, as long as these are included in the information infrastructure Q. The third module of our tool supports the monitoring of already identified topics and developments in the information environment. Considering rules rather than simple sets of terms facilitates a more precise assessment of hangs in the discussion as reflected in the document stream. Moreover, it enables statistical testing for significant changes. The third module is particularly useful for ES activities targeting on competitive intelligence purposes. Therefore, this component might also be suited to the enhancement of other competitive intelligence systems as well.

In summary, our system provides promising evaluation results, but we believe that the main contribution of this paper is the alignment of recent results of psychological studies on ES behavior with state-of-the-art technologies. Consequently, we hope that developers might seize the technologies suggested in this paper to improve their ES systems. Particularly, the cognitive process of ES might point to an auspicious venue of designing improved ES systems.

References

[1] H.A. Simon, "Designing Organizations for an Information-Rich World," Computers, Comm., and the Public Interest, M. Greenberger, ed., pp. 37-52, 1971.

[2] C.-P. Wei and Y.-H. Lee, "Event Detection from Online News Documents for Supporting Environmental Scanning," Decision Support Systems, vol. 36, pp. 385-401, 2004.

[3] R. Decker, R. Wagner, and S. Scholz, "Environmental Scanning in Marketing Planning—An Internet-Based Approach," Marketing Intelligence and Planning, vol. 23, no. 2, pp. 189-199, 2005.

[4] S. Scholz and R. Wagner, "Autonomous Environmental Scanning in the World Wide Web," IT-Enabled Strategic Management: Increasing Returns for the Organization, B. Walters and Z. Tang, eds., pp. 213-242, Hershey: Idea Group, 2006.

[5] M.L. Neugarten, "Seeing and Noticing: An Optical Perspective on Competitive Intelligence," J. Competitive Intelligence and Management, vol. 1, pp. 93-104, 2003.

[6] H. Chen, M. Chau, and D. Zeng, "CI-Spider: A Tool for Competitive Intelligence on the Web," Decision Support Systems, vol. 34, pp. 1-17, 2002.

[7] R.R. Hoffman and J.F. Yates, "Decision(?) Making(?)," IEEE Intelligent Systems, vol. 20, pp. 76-82, 2005.

[8] M. Endsley and R.R. Hoffman, "The Sacagawea Principle," IEEE Intelligent Systems, vol. 17, pp. 80-85, 2002.

[9] J. Holland, K. Holyoak, R. Nisbett, and P. Thagard, Induction: Process of Inference, Learning, and Discovery. MIT Press, 1986.

[10] B. Wierenga, G.V. Bruggen, and R. Stealin, "The Success of Marketing Management Support Systems," Marketing Science, vol. 18, pp. 196-207, 1999.

[11] J. Ontrup and H. Ritter, "Large-Scale Data Exploration with the Hierarchically Growing Hyperbolic SOM," Neural Networks, vol. 19, no. 6, pp. 751-761, 2006.

[12] J. Lamping and R. Rao, "Laying Out and Visualizing Large Trees Using a Hyperbolic Space," Proc. Seventh ACM Symp. User Interface Software and Technology (UIST '94), pp. 13-14, 1994.

[13] P. Pirolli, S.K. Card, and M.M. van der Wege, "Visual Information Foraging in a Focus þ Context Visualization," Proc. ACM Conf. Human Factors in Computing Systems (CHI '01), CHI Letters, vol. 3, no. 1, pp. 506-513, 2001.
[14] H. Ritter, "Self-Organizing Maps in Non-Euclidian Spaces," Kohonen Maps, E. Oja and S. Kaski, eds., pp. 97-110, Amer Elsevier, 1999.
[15] J. Walter, J. Ontrup, D. Wessling, and H. Ritter, "Interactive Visualization and Navigation in Large Data Collections Using the Hyperbolic Space," Proc. Third IEEE Int'l Conf. Data Mining (ICDM '03), Nov. 2003.
[16] H.S.M. Coxeter, Non-Euclidean Geometry. Univ.of Toronto Press, 1957.
[17] T. Kohonen, Self-Organizing Maps, third ed., ser. Springer Series in Information Sciences, 2001.
[18] P. Pirolli and S. Card, "Information Foraging," Psychological Rev., vol. 106, pp. 643-675, 1999.
[19] D.W. Stephens and J.R. Krebs, Foraging Theory. Princeton Univ. Press, 1986.
[20] K. Carley and M. Palmquist, "Extracting, Representing, and Analyzing Mental Models," Social Forces, vol. 34, pp. 921-945, 1995.
[21] J.P. Walsh, "Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane," Organization Science, vol. 6, no. 3, pp. 280-321, 1995.
[22] G.P. Hodgkinson, "Cognitive Inertia in a Turbulent Market: The Case of UK Residential Estate Agents," J. Management Studies, vol. 70, no. 3, pp. 601-636, 1997.
[23] D.S. Kaufer and K.M. Carley, Communication at a Distance: The Effect of Print on Socio-Cultural Organization and Change. Lawrence Erlbaum Assoc., 1993.
[24] J.R. Anderson, The Adaptive Control of Thought. Lawrence Erlbaum, 1990.
[25] S. Scholz and R. Wagner, "The Quality of Prior Information Structure in Business Planning: An Experiment in Environmental Scanning," Proc. Operations Research (OR '04), pp. 238-245, 2005.
[26] W. Boulding, M. Moore, R. Staelin, K. Corfman, P. Dickson, G. Fitzsimons, S. Gupta, D. Lehmann, D. Mitchell, J. Urbany, and B. Weitz, "Understanding Managers' Strategic Decision Making Process," Marketing Letters, vol. 5, no. 4, pp. 413-426, 1994.
[27] B. Liu, Y. Ma, and R. Lee, "Analyzing the Interestingness of Association Rules from the Temporal Dimension," Proc. First IEEE Int'l Conf. Data Mining (ICDM '01), pp. 377-384, 2001.
[28] S. Baron, M. Spiliopoulou, and O. Gu¨ nther, "Efficient Monitoring of Patterns in Data Mining Environments," Proc. Seventh East European Conf. Advances in Databases and Information Systems (ADBIS '03), pp. 253-265, 2003.
[29] C.-H. Lee, M.-S. Chen, and C.-R. Lin, "Progressive Partition Miner: An Efficient Algorithm for Mining General Temporal Association Rules," IEEE Trans. Knowledge and Data Eng., vol. 15, pp. 1004- 1017, 2003.
[30] R. Hilderman and H. Hamilton, "Evaluation of Interestingness Measures for Ranking Discovered Knowledge," Advances in Knowledge Discovery and Data Mining, G. Williams and Q. Li, eds., pp. 247-259, 2001.
[31] S. Brin, R. Motwani, J. Ullman, and S. Tsur, "Dynamic Itemset Counting and Implication Rules for Market Basket Data," Proc. ACM SIGMOD Int'l Conf. Management of Data SIGMOD '97), J. Peckham, ed., pp. 255-264, 1997.
[32] R. Wagner, Mining Promising Qualification Patterns, D. Baier and K. Wernecke, eds., pp. 249-256, Springer, 2005.
[33] S. Kaski, K. Lagus, T. Honkela, and T. Kohonen, "WEBSOM— Self-Organizing Maps of Document Collections," Neurocomputing, vol. 21, pp. 101-117, 1998.
[34] J. Lamping, R. Rao, and P. Pirolli, "A Focus þ Content Technique Based on Hyperbolic Geometry for Viewing Large Hierarchies," Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI '95), pp. 401-408, May 1995.
[35] M. Czerwinski and K. Larson, "The New Web Browsers: They're Cool but Are They Useful," Proc. Seventh Int'l Conf. Human- Computer Interaction (HCI '97), People and Computers XII, 1997.
[36] K. Mullet, C. Fry, and D. Schiano, "On Your Marks, Get Set, Browse!" Proc. ACM Conf. Human Factors in Computing Systems (CHI), 1997.
[37] P. Pirolli, S. Card, and M. van der Wege, "The Effects of Information Scent on Visual Search in the Hyperbolic Tree Browser," ACM Trans. Computer-Human Interaction (TOCHI '03), vol. 10, no. 1, pp. 20-53, 2003.