Improving community-portals usability

C. F. Baizan, S. Eibe, and O. Marban

Facultad de Informática, UPM. Madrid, Spain
{cfbaizan, seibe, omarban}@fi.upm.es

Abstract. Humans and machines are not using today the same formats of knowledge representation. This gap makes a serious problem in the interaction human-machine. We are exploding intelligent methods to minimize this gap showing at different human users, specifically and personalized, the knowledge used by machines.

1 Introduction

The goal of the emerging Second-Generation Web or Semantic Web is to mark-up the meaning of Web data to make them understandable and automatically processable by computers. So, many Semantic Web initiatives emphasize the ability for exchanging the meaning of information. Although these efforts will lead to an improved quality of the application results, user interfaces often doesn't draw profit from the semantics gain.

But from the user's perspective, what people says when they are asked about what they consider as the most important problems using the web, most answers indicate that finding new information is the main problem. Even to go back to pages already visited is difficult. So, we need to allow people to find information both in a fast and pleasant way.

Anyway, providing explicit and formal descriptions of meaning (metadata) in relation with data structures, process, devices, networks, ... but also for human-consumption. There are several emerging initiatives (such as the Dublin Core Metadata Initiative [15]) engaged on developing interoperable online metadata standards, but all of them are only for machine-use. In fact, machines are never (or in the near future, at least) going to understand knowledge under those descriptions. Their interpretation is reserved to humans. So, humans and machines are not using today the same formats of knowledge representation. On the one hand, analysis of unconstrained natural language is very difficult for machines and on the other hand neither RDF nor XML are readable by humans. This gap is a serious problem in the interaction human-machine.

Both humans and machines need to look for some common representation of knowledge. Visual Semantic Web (VSW) is a new paradigm where this problems are considered. VSW is an enhancement of the current Semantic Web, in the same way as Semantic Web improves traditional Web. As a consequence of this evolutionary design, a first question is if humans and machines must use different

formats of knowledge representation or an integrated view could be possible. We are now exploring methods to minimize this gap, by showing to different human users the knowledge representation used by machines.

In order to fill this gap, enterprises have been worried about identifying the navigator. What is important is not the identity of the user but information about his likes and dislikes, his preferences, the way he behaves. All this, integrated with information related to the activity, will result in a warm human relationship with users. The relationship with users is paramount when aiming to develop activities in any web environment trying to keep the traditional feeling of the user. So adapting the web site to user preferences is the unavoidable commitment that web-site sponsors must face.

Obtaining and examining the non explicit but available knowledge about customers and sponsors is the way to follow to have success in any web activity. To analyze the available user navigation data and in order to obtain from them knowledge, web mining techniques have to be used.

The rest of the paper is organized as follows. First, we present summarized related work about Knowledge Management Tools and the Semantic Web and then about related Web Mining. Next, we present our vision about the web usability problem solving approach combining the visualization of semantics and web mining. Finally, we conclude by giving a few directions about our nearest future work.

2 Related work

2.1 Knowledge Management Tools and Semantic Web

The Semantic Web seems to be a solution to the problem of Knowledge Management in the Web. Merz [31], one of the first proposals about knowledge management, is a tool for personal information management and knowledge work that addresses usability problems like information overload, disorientation and difficulties in organizing information. Merz proposal is based on personalized metaphors to information visualization with hand-written only (not automatic) annotations. More recently, a lot of work has been done about web knowledge management, specially on ontologies and the semantic web. The Ontobroker project [13] uses ontologies to annotate and wrap Web documents and provides an ontology-based answering service OntoBroker.

The problem of organizing and presenting knowledge is a part of the more general one of knowledge management. Concept maps [25], Topic Maps [43, 37, 38] and Kohonen's Self-Organizing Maps (SOM) [28, 29] are tools for organizing and representing knowledge. The WEBSOM [27] is a specific SOM-based method that have been successfully applied to web documents [46] and textual data mining [30]. More recently several works have combined SOM with topics and categories [45].

Metadata are descriptions associated with a given ontology. Annotation [20] is the process of creating and associating metadata with web pages. These pages

usually exist previously but they also can be generated dynamically from a database [22]. Hence, our interest is focused on the visualization of semantics, specially the one associated with relations between concepts. Such a metadata is named in *CREAM* [21] *relational metadata* to denote those annotations describe relationships between class instances.

Information Visualization [9] is a new and rapidly growing research area which aims to provide visual depictions of very large information spaces. It covers interdisciplinary areas such as Information Retrieval [4], World Wide Web [33] or Human Computer Interaction [17]. One of the most active and promising research topic in this field is visualization of semantics. The starting point is to look for descriptions of contents instead of for true contents. It is often easier to process metadata automatically than ordinary content. It is less ambiguous and also it allows an easy computational description. [19, 18, 44] show methods to support user tasks through visualization and uses ontology-based visualization methods to data analysis, querying and navigation. In [35] Conzilla, a first prototype of a concept browser, is introduced. Conzilla is a tool that creates an overview of the different concepts under study by supporting a separation between context and content.

Viewing the Web as a large knowledge base, we get the reasons to visualize the web. Particularly, in a search-centric view of web finding never-before encountered information and extracting knowledge from the web [11] assets a most specific goal where visualization can help us. Therefore, visualizing the Worl Wide Web [33] requires new front-end tools to aid navigation and search interfaces [41] that follow the structure of information. In [3] tools to navigate through a set of documents that are clustered for different user needs and where each cluster is labelled with related words are presented. The Atlas of Cyberspace [14] introduces an atlas showing the best maps and graphic representations of the Internet, WWW and Cyberspaces. Cat-a-Cone [23] introduces a novel user interface that integrates search and browsing of very large category hierarchies with their associated text collections. An essential component is the separate but simultaneous display of the representations of the categories and the retrieved documents. Another main component is the display of multiple categories simultaneously selected, together with their hierarchical context. In order to improve the organization of search results, [16] proposes to show *hierarchical faceted metadata* in an appealing and understandable way. Faceted metadata are defined as orthogonal descriptors within a metadata system that are used to organize the structure of a web site and search results.

2.2 Web Mining

Web mining is a broad term that has been used to name different processes: information discovery from sources on the web (web content), discovery of the structure of the web servers (web structure) and mining user browsing and access patterns through logs analysis (web usage) [1]. In particular, research in web usage has focused on discovering access patterns from log files.

Web servers record data about user interactions. The analysis of web access logs can help to understand the user behaviour and the web structure. From the business point of view, knowledge obtained from the Web usage patterns could be directly applied to efficiently manage activities related to e-services, because the relationship with the customer is paramount in current web environment. In order to analyze the logs, different web mining methods have been proposed. The user behaviour could be understood by combining users navigation paths with other data features, such as page viewing time, hyperlink, and page content [24].

Based on data mining techniques, many approaches [7, 39] have been proposed to extract frequent paths in order to understand user behaviour through data mining techniques. An hybrid approach [40] to analyze the visitor click sequences is proposed. A combination of hypertext probabilistic grammar and click fact table approach is used to mine web logs which could be also used for general sequence mining tasks. Mobasher [34] proposed the web personalization system which consists of offline tasks related to the mining of usage data and online process of automatic Web page customization based on the discovered knowledge. LumberJack [10] builds up user profiles by combining user session clustering and statistical traffic analysis using K-means algorithm.

Consequently, the challenge is to get, analyze and understand the behavior of every user that connects to the web site. Based on annotations and ontologies, several methods and approaches have been proposed in order to take these site semantics into account [6, 12, 42].

3 Usability improvement in the Visual Semantic Web

Our proposal to improve Web usability is to build web interfaces by combining data and treatment exhibition on a per-site and per-user basis. Navigation and information retrieval on a Web site are not easy tasks because users have to extract information from the large amount of data available. Moreover, most of these data are unstructured, which complicates the application of existing data mining techniques. However, the join of new semantic structures to the Web Mining process will improve the results in several ways:

- It will dynamically provide specialized interfaces, by combining both textual and non textual metaphors. Double-clicking on a metaphor object gives access to various actions: exploding or comprising, annotating, or even removing it.
- It will allow the dynamic organization of the web site structure based on metadata about users (per-user) and contents (per-site). Different users will require different views of information so the metaphors presenting contents will be per-user generated.
- It will provide means of visualization of the semantics (metadata and so on) related to the organization of the site contents.
- Depending on the scale to which the content is viewed it can be represented by different objects (title, summary, ...). Users could choose a partial view

and need to know the available filters to select the most semantically significant contents. Therefore, filter descriptions will be needed.
- It will support interactive browsing of contents by means of an adequate visualization of metadata about categories and relations.
- For visualization of semantics purposes, basic metadata are taken from well known sets as Dublin Core, but for the visualization of more detailed descriptions of contents, metadata are directly taken from the interaction with the user (user defined categories) or automatically derived (semi-automatic metadata creation). Users can interactively define new descriptions and notes about contents.

In the following, we are going to explain the topics that have been previously introduced.

3.1 Metaphors to Knowledge Visualization

The first question is that, in order to be able to exchange the semantics of information being distributed and shared among many users, it is first necessary to agree on how to explicitly model it. Ontologies are a tool for representing such formal and shared domain descriptions. Although, in general, ontologies are not oriented to visualization, the majority of the ontologies in community portals are *light-weight* [19], i.e., they are predominantly taxonomies, with very few cross-taxonomical links and with relatively very few logical relations between the classes. Moreover, the number of instances will typically be very large in comparison with the number of classes. Community portals are paradigmatic because the information contents that they offer are about the various kinds of work in the organization. Therefore, ontology visualization can help to exhibit data and treatment of this knowledge as humans seem understand it (semantics). For example, a web search interface to an online university community portal could use an ontology-based navigational search engine when composing, evaluating and displaying the user's query result. At the query formulation stage it could use the ontology to assist the user with meaning about organization and properties of contents, but it could also use it to improve the presentation of the resulting list to the final user, in a personalized way, by replacing the endless list of hits with a navigation and browsing structure based on the semantics of the hits. Besides ontologies there are another context-based ways of describing information semantics. Concepts belonging to the same context share a set of common but local properties. As a result, the meaning of related domain concepts can be fixed without vagueness and ambiguity by context visualization. Additionally, contexts are best used in applications as classifications of documents, the development and integration of catalogs [8] and in presenting domain specialized informational content.

From a more pragmatical point of view, presenting content requires the use of metaphors to bind user knowledge (pre-attentive processing). Metaphors are similar to "containers" for the information that is to be presented. Roughly speaking, the goal of metaphors is presenting contents and their right context.

In the case of a hyper-linked system, concepts and metaphors representing them generally appear in many different contexts, whose number and form are constantly changing by the addition and removal of pages and links. This makes it hard to maintain a clear separation between context and content. Consequently, users often asked for the context corresponding. In fact, actually a web page may be see as a container of its content as well as a context for the contents that are reachable (by a mouse click) from it. This underlying link structure leads to an inextricable mixture of context and content where the conceptual relationships of the content are lost.

Therefore, separation between context and content is necessary and this is only possible with metaphors enhanced by semantic data. In this way, content-oriented category metaphors provides semantics visualization support. These kind of metaphors will be used to build search and browsing specialized interfaces that explicitly support the types of tasks users perform at each specific site, i.e., in a per-site basis. Moreover, as searching is the primary usability problem, we are looking for integration of search and analysis information. This requires a kind of intelligence to support series or cycles of searches (iteration and refinement), backing up and moving forward, next step suggestion and comparison of resulting data. Briefly, it requires context and (user) interactivity support to present metadata that organizes contents and comment search results.

Although many attention has been paid to how create metadata, very little has been paid to how explicitly use them in web user interfaces. So, a main challenge is the choice and design of the the most adequate metaphors with their components and properties. A good metaphor [5] must not only distinguish context from content, but it also must support user interaction. Moreover, visualization metaphors must guide user learning. Usually, users learn during the search process while navigating the site or, even worst, according to a trial and error navigation pattern. So, interfaces should make it easy to store intermediate results and to follow trails with unanticipated results although it makes evaluation more difficult. For example, a zoomable tree-based history map of recently visited web-pages, where user can shrink sub-hierarchies placed to left of browser window, contributes to user navigation by portal contents. Also, when the user makes a query, such a tree can be dynamically combined with search results to serve as a table of query related contents at the site. Moreover, descriptions of contents (i.e., category and subcategory labels) will be picked up according to manually and machine-created categories based on information extraction about the behavior of other users.

As we previously said a main component of the interface is the separation of the graphical representation of the category hierarchy from the graphical representation of the contents. This separation allows for a fluid and flexible interaction between browsing and searching, both over categories and contents. It also provides a tool by means of which a set of categories can be related with their hierarchical context (contents are reached from different category labels), reflecting what has people to do with the site. In contrast to previously existing systems as Letizia [32] or WebWatcher [26], in which the strategy for giving

advice is learned from feedback from earlier tours and based on user profile, our approximation focused on information delivery based on visual explanations rather than on improving performance.

3.2 Semantics and Web Mining

We propose web mining on data and metadata to present information to the user according to his preferences. So the first step is to identify the user preferences.

We assume that the server has running a log system which can annotate the log with the appropriate semantic according to an ontology. This system can be similar to the one introduced in [36] (see figure 1.a). In this paper, we propose how we can improve the portal web usability through intelligent visualization techniques.

First of all, it is necessary to know about the surfers' preferences in order to display the information according to those preferences. But the information has to be organized in an user-friendly and comprehensive way that can be located straightforwardly. We propose to use metaphors to achieve that.

We suggest to use a tree as a metaphor. The tree will show the searched information to the user according to previous surfers navigation paths leading to the same information. In order to show that information, we need to obtain the frequent paths traversed by the previous users. This can be done by using some of the algorithms that calculated the frequent paths (see section 2.2) according to the information recorded in the web log file.

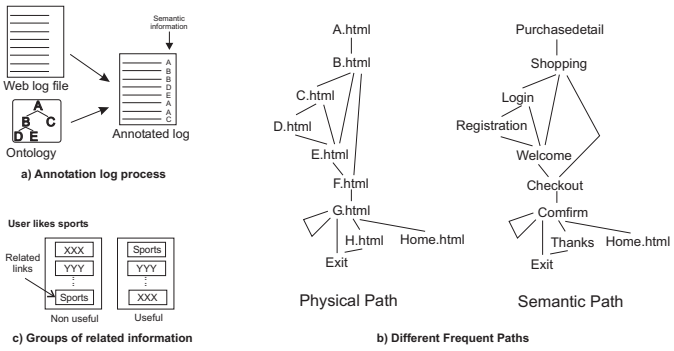

Fig. 1. Different kinds of items in the visualization process

Inasmuch as we can obtain an annotated log with semantic information, the semantic information can be used to complete or to modify the frequent paths' information, so we can obtain different kind of frequent paths. On the one hand, we can obtain frequent paths which represent links between web pages and how users traversed them. Also, we can obtain semantic frequent paths. These semantic frequent paths can be obtained if we only show the semantic information that is associated in the semantic web log. But we can have a mixed

frequent path, with navigational and semantic information, in order to improve the user navigation process (see figure 1.b).

When the information is displayed to the user, it can be aggregated through the defined ontology used to annotate the log. The aggregated information is useful for specializing and generalizing the showed concepts. So, the user can begin to look for generic information and from that "generic" information can traverse to more specific information. We have to use the same ontology that was used to annotate the log, to allow to go from a concept to a more specific or generic one through the appropriate reasoning ontology method.

The creation of information clusters to show to the user the related information in groups is another improvement. It is necessary to apply clustering data mining techniques over the semantic of the web pages and over the users' profiles to obtain clusters of web pages that were visited by the same kind of user. So, when a new user will come to visit the site, we can establish a profile for him/her. Once the profile is established, information of highest value to the user can be showed without the intervention of the user, according to the calculated preferences of the user.

In order to improve the navigational process, we can show to the user information related to the target one (this is the way in which Amazon (www.amazon.com) and other organizations are working). Additional information related to the target information is displayed when the target information is showed. In order to implement this mechanism, we can apply data mining techniques. The most adequate technique to do it, is association rules, applying, for example Apriori algorithm [2]. This algorithm calculates which pages are visited in the same session. Thus, when the user looks for some information, we display the results of his/her search but we also display new links. Those links were traversed by other users in their navigation, so the new user can go directly to the new information without the necessity of traverse the physical path through the site.

To improve the navigation through the site, a very useful improvement could be the annotation of links. Thus, the user can put some label to the links in order to have personal information of the links. This can be useful to reproduce some searches in the site or to go back to the right previous link.

Finally, the information in the web page must be organized according to the user preferences, for example, if the user likes sports, when he/she performs a search, the result must be displayed in the right order, first the sport information and then the rest of it (see figure 1.c).

4 Conclusions

Our work is exploring the advantages of semantics visualization in community portals. In this paper we present a way to extend the application of intelligent methods to display to different human users, specifically and personalized, the knowledge used by machines. Therefore, we consider the improvement of visualization metaphors incorporating semantic (visualizing metadata), which is obtained manually as well as automatically, applying web mining techniques.

References

1. A. Abraham. I-miner: A web usage mining framework using hierarchical intelligent systems. In *The IEEE Int. Conf. FUZZ-IEEE'03*, pages 1129–1134, 2003.
2. R. Agrawal and R. Strikant. Mining sequential patterns. In *Proc. 1995 Int. Conf. Very Large Databases*, pages 487–499, 1994.
3. P. Au, M. Carey, S. Sewraz, Y. Guo, and S.M. Rüger. New paradigms in information visualization. In *Research and Development in Information Retrieval*, 2000.
4. R. Baeza-Yates and B. Ribeiro. *Modern Information Retrieval*. ACM Pressç, 1999.
5. P. Barr, R. Biddle, and J. Noble. A taxonomie of user-interface metaphors. In *Proc. of Fourth Australasian User Interface Conf.*, February 2003.
6. B. Berendt. Using site semantic to analyze, visualize, and support navigation. *Data Mining and Knowledge Discovery*, 6:37–59, 2002.
7. J. Borges and M. Levene. A heuristic to capture longer user web navigation patterns. In *Proc. Of the 1st Int. Conf. On Electronic Commerce and Web Technologies*, Greenwich, U.K., September 2000.
8. Paolo Bouquet, F. Giunchiglia, F. van Harmelen, Serafini, and H. Stuckenschmidt. C-OWL: Contextualizing Ontologies. In *Procs Second International Semantic Web Conference. ITC-IRST Technical Report #0306-06*, May 2003.
9. S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision To Think*. Morgan Kaufmann, 1999.
10. E.H. Chi, A. Rosien, and J. Heer. LumberJack: Intelligent discovery and analysis of web user traffic composition. In *Proc.Of ACMSIGKDD Workshop on Web Mining for Usage Patterns and User Profiles*, 2002.
11. M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slattery. Learning to extract symbolic knowledge from the world wide web. In *Proc. of the 15th National Conf. on Artificial Intelligence (AAAI-98)*, 1998.
12. H. Dai and B. Mobasher. A road map to more effective web personalization: Integrating domain knowledge with web usage mining. In *Proc.of the International Conference on Internet Computing 2003 (IC'03), Las Vegas, Nevada*, June 2003.
13. S. Decker, M. Erdmann, D. Fensel, and R. Studer. Ontobroker: Ontology based access to distributed and semi-structured unformation. In *DS-8*, page ç, 1999.
14. M. Dodge and R. Kitchin. *The Atlas of Cyberspace*. Addison-Wesley, 2002.
15. Dublin Core Metadata Initiative. *Dublin Core Bibliography*. http://dublincore.org/documents/usageguide/bibliography.shtml.
16. J. English, M. Hearst, R. Sinha, K. Swearingen, and P. Yee. Hierarchical faceted metadata in site search interfaces. In *Proc. Of ACM CHI 02*, 2002.
17. A. Dix et al. *Human-Computer Interaction, 2nd Edition*. Prentice Hall, 1998.
18. C. Fluit, M. Sabou, and F. van Harmelen. Ontology-based information visualisation. In *Visualising the Semantic Web*. Springer Verlag, 2002.
19. C. Fluit, M. Sabou, and F. van Harmelen. Supporting user tasks through visualisation of light-weight ontologies. In *Handbook on Ontologies in Information Systems*. Springer-Verlag, 2003.
20. S. Handschuh and S. Staab. *Annotation for the Semantic Web. Volume 96 Frontiers in Artificial Intelligence and Applications*. IOS Press, 2003.
21. S. Handschuh and S. Staab. CREAM - CREAting metadata for the semantic web. *Computer Networks, Elsevier. Special Issue on the Semantic Web*, 2003.
22. S. Handschuh, S. Staab, and R. Volz. On deep annotation. In *Proc. of the 12th int. conf. on World Wide Web*, pages 431–438. ACM Press, 2003.

23. M.A. Hearst and C. Karadi. Cat-a-Cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. In *Proc. Of SIGIR-97*, pages 246–255, Philadelphia, US, 1997.
24. J. Heer and E. H. Chi. Identification of web user traffic composition using multi-modal clustering and information scent. In *Proc. Of the SIAM*, pages 51–58, 2001.
25. J. D. Novak, Cornell University. *The Theory Underlying Concept Maps and How To Construct Them.* http://cmap.coginst.uwf.edu/info/index.html.
26. T. Joachims, D. Freitag, and T. Mitchell. Webwatcher: A tour guide for the world wide web. In *15th Int. Joint Conf. IJCAI-97*, Nagoya, Aichi, Japan, August 1997.
27. S. Kaski, T. Honkela, K. Lagus, and T. Kohonen. Websom—self-organizing maps of document collections. *Neurocomputing*, 21(1):101–117, 1998.
28. T. Kohonen. The self-organizing map. In *Proc. of the IEEE*, volume 78, pages 1464–1480, 1990.
29. T. Kohonen. A look into the self-organizing maps. In *Proc. Of the ICSC Symposia on Neural Computation (NC'2000)*. ICSC Academic Press, 2000.
30. K. Lagus, T. Honkela, S. Kaski, and T. Kohonen. WEBSOM for textual data mining. *Artificial-Intelligence-Review*, 13:345–64, 1999.
31. S. Lenman, H. See, M. Century, and B. Pennycook. Merz: Creating personal and shared spaces on the worldwideweb, 1996.
32. H. Lieberman. Letizia: An agent that assists web browsing. In *Proc. of the Int. Joint Conf. on Artificial Intelligence*, Montreal, August 1995.
33. A. Mendelzon. Visualizing the World Wide Web. *White Paper. University of Toronto*, 1996.
34. B. Mobasher, H. Dai, T. Luo, M. Nakagawa, and J. Witshire. Discovery of aggregate usage profiles for web personalization. In *Procs. Of the WebKDD*, 2000.
35. A. Naeve. The concept browser - a new form of knowledge management tool, 2001.
36. D. Oberle, B. Berendt, A. Hotho, and J. Gonzalez. Conceptual user tracking. *Advance in Web Intelligence. AWIC 2003*, pages 155–164, May 2003.
37. Ontopia. *Ontopia: Learn more about topic maps. http : //www.ontopia.net/topicmaps/learn_more.html.*
38. Jack Park and Sam Hunting. *XML Topic Maps: Creating and Using Topic Maps for the Web.* Addison-Wesley, 2002.
39. J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu. Mining access patterns efficiently from web logs. In *Proceedings of the PAKDD'00*, 2000.
40. S.E. Jespersen S.E, J. Thorhauge, and T. Bach. A hybrid approach to web usage mining. In *Data Warehousing and Knowledge Discovery*, pages 73–82, 2002.
41. M.M. Sebrechts, J. Cugini, S.J. Laskowski, J. Vasilakis, and M.S. Miller. Visualization of search results: A comparative evaluation of text, 2d, and 3d interfaces. In *Research and Development in Information Retrieval*, pages 3–10, 1999.
42. M. Spiliopoulou and C. Pohle. Data mining for measuring and improving the success of web sites. *Data mining and Knowledge Discovery*, 5(1-2), 2001.
43. Steve Pepper. *The TAO of Topic Maps.* http://www.ontopia.net/topicmaps/materials/tao.html.
44. F. van Harmelen, J. Broekstra, C. Fluit, H. ter Horst, A. Kampman, J. van der Meer, and M. Sabou. Ontology-based information visualisation. In *Proc. of the workshop on Visualisation of the Semantic Web (VSW'01)*, September 2001.
45. C. C. Yang, C. Hsinchun, and K. Hong. Internet browsing: visualizing category map by fisheye and fractal views. In *Proc. Int. Conf. On Information Technology: Coding and Computing*, pages 34–9, 2002.
46. Christopher C. Yang, Hsinchun Chen, and Kay Hong. Visualization of large category map for internet browsing. *Decis. Support Syst.*, 35(1):89–102, 2003.