Applying Dynamic Causal Mining in Retailing

Yi Wang

Abstract—**With the fast development of information technology, retailers are suffering from the excess of information. Too much information can be a problem. However, more information creates more opportunity. In retailing, information is the key issue to maximizing revenue. It is now hard to make timely or effective decisions and to the right content to the right place, at the right time and in the right form. This paper is about managing the information so that the user can gain more clear insight. It is about integrating and inventing methods and techniques. The Semantic Web will provide a foundation for such a solution. However, semantics only provide a way of mapping the content of a web to user defined annotations. Not many companies have fully utilized the power of Internet retailing due to the various technical obstacles have yet to be overcome. The existing research in e-retailing focuses only on the traditional retailing including direct and indirect retailing approaches. This paper suggests that applying association mining techniques can further improve the dealing of information overload in a web oriented retailing environment.**

Index terms—**Semantic web, online retailing, data mining, formal concept, Protégé, triple store, Sparql.**

I. INTRODUCTION

Information is all around us, easy to collect, store and access. It consists of the useful data that is needed to solve our problems. But as information is more and more overloaded, managers, researchers or retailers have to spend more time to process the information before making their decisions, due to the fact that the stored information is unstructured.

One way of solving this problem is to turn the important and useful information into knowledge and filter away the less important information. This requires understanding of when to use information, how to find it, and how to present it to the target customer. It is imperatives that enterprises will need to exploit the knowledge and information available on the WWW as far as possible, so its static nature will tend to increase the information overload. This paper suggests using Semantic Web to solve the information overload problem. The technological basis of the Semantic Web provides a unified framework within which many approaches to the problem of information overload can be integrated.

The major components in Semantic Web are shared conceptualizations and terminologies to describe customers, operation, content of web page, etc. These conceptualizations and terminologies are called ontologies. They may refer to agreed ways of describing customer's preference and demand, operational capacity and constraints, retailing brand. Ontologies can be used as meaningful enrichment for the content. They provide common base framework within which information can be properly shared, modeled and filtered. The next part of paper reviews some existing literature regarding online retailing and information overload. The third part of the paper explains semantic web as a mean to solve information overload. The fourth part gives some technical background for the usage of semantic web in retailing. The fifth part suggests some improvement based on existing technologies. The sixth part will discuss in detail about dynamic causal mining as a specific data mining tool. The seventh part will illustrate a practical problem. The final part concludes the whole paper.

II. LITERATURE REVIEW

On-line retailing offers more choices and flexibility [Lamoureux, 97] and, at the same time, eliminates huge inventories, storage costs, utilities, space rental, etc. [Avery, 97]. Companies can design and personalize advertising for each customer [Peterson, *et al*., 1997]. The Internet can provide timely information to customers because of its ability for instant communication [Lane, 1996]. This means more interaction [Rosenspan, 2001] and quicker responses [Isaac, 1998]. The information can be used to assist new product development and introduction [Higgins, 2001]. The communication also helps with identifying prospects [Ebling, 2001], sales and relationship building [Ginovsky, 2001], and deepening customer loyalty [Kranzley, 2001]. Perlow [1999] describes a software company characterized by an environment. It also allows for easy follow-up on customers' needs [Marks, 1998]. Retailing activity occurs through three types of channels: communication, transaction, and distribution channels [Peterson, 1997].

Studies of the semantics web were initiated by Tim Berners-Lee, the creator of the World Wide Web [Berners-Lee, 01]. The Web is referred to as the "semantic Web", where information will be retrieve from intelligent network services such as information brokers and search agents [Decker & Melnik, 00, Decker *et al*., 00].

The World Wide Web has evolved into a dynamic, distributed, heterogeneous, complex network, which is hard to

control [Albert *et al*., 99, Huberman & Adamic, 99]. It is important to have consistent understanding and interpretation [Helbing *et al.*, 00,]. in the World Wide Web [Huberman & Adamic, 99, Huberman *et al.*, 97, Barabasi & Albert, 99].

When more information arrives than individuals can process, an information overload occurs [Simon, 1971]. Much research has done in dealing with information overload. Abiteboul *et al.* [Abiteboul, 00] systematically investigated the data on the Web and the features of semistructured data. Zhong studied text mining on the Web including automatic construction of ontology and filtering system [Zhong, 00; Zhong *et al.,* 00b]. Liu *et al*. worked on e-commerce agents [Liu & Ye, 01] and KDDA (Knowledge Discovery and Data Mining Agents) [Liu & Zhong, 99, Liu *et al*, 01] to minimize the information overload.

III. A GENERAL FRAMEWORK OF SEMANTIC WEB

Figure 1 shows the simplified representation of a semantic network where enter prise on one side provides information and the user on the other hands give the queries.
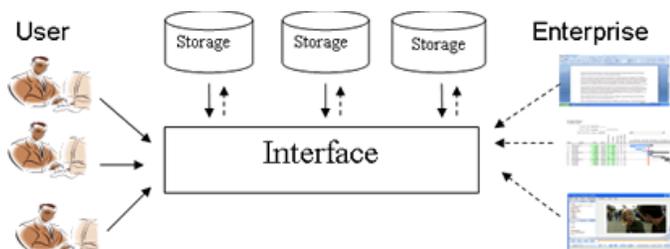

Figure 1 A general frame work

The semantic network interface has three goals:
1. **To provide ontologies for interoperations.** In many cases, the customer and the enterprise does not speak the same language. They have different preference, gain different knowledge and get different information from the product or services. In order to have a more successful relationship, there is need for some ontology which bridges these gaps.
2. **To unify information from different document format**. The enterprises provides the information online in different format, some as word documents, some as excel documents, some as media files., etc. The goal here is to integrate all relevant based on the given ontology. And map all the information to a user friendly representation.
3. **To Store the relevant information and update the ontology.** The interface should be able to retrieve and represent the information based on user's queries. And the interface should be able to update the ontologies for improvement, thus rather storing a large amount of information, the relevant ontologies or relations are stored.

IV. ONTOLOGY DEVELOPMENT AND STORAGE

An ontology for retailing defines a common vocabulary for any participants in retailing, including customer, manager, retailer, etc. who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them. The major goals for ontology development (Natalya and McGuinness, 07) are:
1. To share common understanding of information among user (Musen 1992; Gruber 1993).
2. To enable reuse of domain knowledge was one of the driving forces behind recent surge in ontology research.
3. To change domain assumptions if information about the domain changes.
4. To separate the domain information from the operational information
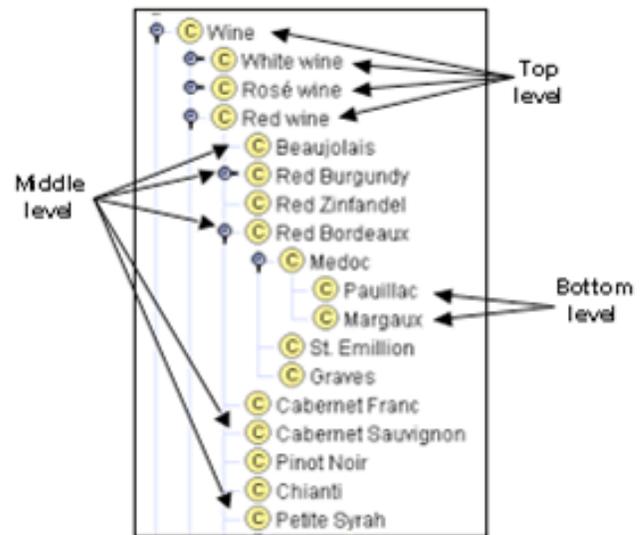5. To reuse the existing ontologies and extending them (McGuinness *et al*. 2000).

Figure 2 Wine brand taken from(Natalya and McGuinness, 07)

The most common tools for developing ontology are Protégé-2000 (Protege, 2000), Ontolingua (Ontolingua, 1997), and Chimaera (Chimaera, 2000) as ontology-editing environments.

Figure 2 shows a sample of the protégé interface for wine ontology. It is typically based on graphical class hierarchical development (Uschold and Gruninger, 96). An ontology can be seen as a triplets (Subject, relation, object). An typical example is "Wine", which is subject, "has brand name", which is a relation, "Chinati" which is an object. Instead of normal Sql technology, the ontology can be store in a triple store, which is a specific type of data storage. It is designed to store and retrieve identities that are constructed from triplex collections of *strings* (sequences of letters) and can be queried with Sparql. Sparql query consists of three parts. The pattern matching part, which focus on matching patterns of graphs, like optional parts, union of patterns, nesting, The solution modifiers part , which allows to modify values applying classical operators like projection, distinct, order, limit, and offset. And the output part which consists of different types: yes/no solution based on descriptions of resources.

V. MINING TECHNIQUES

The ontologies developed for semantic web are based on user experience. This requires the need for the developers to have an increased understanding of the complex issues involved in the ontology. And sometimes it is difficult to make a universal accepted ontology. And a lot of hidden relations are not modeled at all. This section suggests combining classical ontology development with data mining for identifying hidden information and expanding the application area of both techniques. This gives an improved description of the target system represented by a database; it can also improve strategy selection and other forms of decision making.

Data mining techniques to automatically discover and extract information from Web documents and services [Kosala & Blockeel, 00, Srivastava *et al.*, 00, Zhong, 01]. Zhong *et al.* proposed a way of mining peculiar data and peculiarity rules that can be used for Web-log mining [Zhong, 99]. They also proposed ways for targeted retailing by mining classification rules and retail value functions [Yao and Zhong, 01, Zhong *et al.*, 00]. Data mining is the systematic refining of information resources on the Web for business intelligence [Hackathorn, 00].

This paper suggests using associative formal concept analysis as the base tools for data mining (Wolff, 94) and develops it further using Dynamic causal mining as the technique for mining the relations. The DCM algorithm was discovered in 2005 [Pham *et al.*, 2005] using only counting algorithm to integrate with Game theory. It was extended in 2006 [Pham *et al.*, 2006] with delay and feedback analysis, and was further improved for the analysis in Game theory with Formal Concept analysis [Wang, 2007]. DCM enables the generation of dynamic causal rules from data sets by integrating the concepts of Systems Thinking [Senge *et al.*, 1994] and System dynamics [Forrester, 1961] with Association mining [Agrewal *et al.*, 1996]. The algorithm can process data sets with both categorical and numerical attributes. Compared with other Association mining algorithms, DCM rule sets are smaller and more dynamically focused. The pruning is carried out based on polarities. This reduces the size of the pruned data set and still maintains the accuracy of the generated rule sets. The rules extracted can be joined to create dynamic policy, which can be simulated through software for future decision making. The rest of this section gives a brief review of Association mining and System Dynamics.

Association mining was discovered by Agrawal [Agrawal *et al.*, 1996]. It was further improved in various ways, such as in speed [Agrawal *et al.*, 1996 and Cheung *et al.*, 1996] and with parallelism [Zaki *et al.*, 1997] to find interesting associations and/or correlation relationships among large sets of data items. It shows attributes value conditions that occur frequently together in a given dataset. It generates the candidate itemsets by joining the large itemsets of the previous pass and deleting those subsets which are small in the previous pass without considering the transactions in the database. By only considering large itemsets of the previous pass, the number of candidate large itemsets is significantly reduced.

Systems thinking is about the interrelated actions which provide a conceptual framework or a body of knowledge that makes the pattern clearer [Senge *et al.*, 1994]. It is a combination of many theories such as soft systems approach and system theory [Coyle, 1996]. Systems thinking seeks to explore things as wholes, through patterns of interrelated actions. *System dynamics* [Sterman 1994 & Coyle, 1996], is a tool to visualize and understand such patterns of dynamic complexity, which is build up from a set of system archetypes based on principles in *System thinking* [Sterman, 2000]. *System dynamics* visualizes complex systems through causal loop diagrams. A causal loop diagram consists of a few basic shapes that together describe the action modeled.

System dynamics addresses two types of behavior, *sympathetic* and *antipathetic* [Pham *et al.*, 2005]. *Sympathetic behavior* indicates an initial quantity of a target attributes starts to grow, and the rate of growth increases. *Antipathetic behavior* indicates an initial quantity of a target attributes starts either above or below a goal level and over time moves toward the goal.

VI. MINING ALGORITHM

A. *Problem Formulation*

Let D denote a database which contains a set of n records with attributes {A1, A2, A3,…, An.}, where each attribute is of a unique type (sale price, production quantity, inventory volume, etc). Each attribute is linked to a time stamp t. To apply DCM, the records are arranged in a temporal sequence (t = 1, 2,…, n). The causality between attributes in D can be identified by examining the polarities of corresponding changes in attribute values. Let D_{new} be a new data set constructed from D. A generalized dynamic association rule is an implication of the form A1 $\rightarrow$p A2, where A1 $\subset$ D, A1 $\subset$ D, A1$\cap$ A2= ϕ and p is the polarity.

The implementation of the *DCM* algorithm must support the following operations:
(1) To add new attributes.
(2) To maintain a counter for each polarity with respect to every dynamic value set. While making a pass, one dynamic set is read at a time and the polarity count of candidates supported by the dynamic sets is incremented. The counting process must be very fast as it is the bottleneck of the whole process.

B. *Algorithm Description*

DCM makes two passes over the data as shown in Figure 3. In the first pass, the *support* of individual attributes is counted and the frequent attributes are determined. The dynamic values are used for generating new potentially frequent sets and the actual *support* of these sets is counted during the pass over the data. In subsequent passes, the algorithm initializes with dynamic value sets based on dynamic values found to be frequent in the previous pass. After the second of the passes, the *causal rules* are determined and they become the candidates for the dynamic policy. In the *DCM* process, the main goal is to find the strong dynamic causal rule in order to form a policy. It also represents a filtering process that prunes away static attributes, which reduces the size of the data set for further mining.

Part 1: – Preprocessing: Removal of the "least" causal data from database
Part 2: – Mining: Formation of a rule set that covers all training examples with minimum number of rules
Part 3: – Checking: Check if an attribute pair is self contradicting (sympathetic and antipathetic at the same time)
Input: The original database, the values of the pruning threshold for the neutral, sympathetic and antipathetic supports.
Output: Dynamic sets
Step 1: Check the nature of the attributes in the original database (numerical or categorical). Initialize a new database with dynamic attributes based on the attributes and time stamps from original database.
Step 2: Initialize a counter for each of the three polarities.
Step3: Prune away all the dynamic attributes with supports above the input thresholds.
Step 4. Check weather a rule is self-contradictory (a rule is both sympathetic and antipathetic).
Step 5. If step 1 returns true then
Retrieve the attribute pair form the preprocessed database
Step 6. Initialize a counter that includes polarity combination
Step 7. For the pair of attributes
Count the occurrence of polarity combination with two records each time.
Prune away the pairs if the counted support is below the input threshold.

Figure 3. The Steps of DCM

Table I
Pruned results

Data set	Single Support						
	0.05	0.10	0.15	0.20	0.25	0.30	0.35
Adult	5%	20%	27%	74%	100%	100%	100%
Bank	11%	20%	60%	94%	100%	100%	100%
Cystine	5%	33%	70%	100%	100%	100%	100%
Market basket	6%	10%	50%	86%	100%	100%	100%
Mclosom	1%	13%	38%	72%	90%	100%	100%
ASW	1%	8%	40%	68%	70%	97%	100%
Weka-base	6%	25%	52%	86%	100%	100%	100%

VII. EXPERIMENT

A. Data preparation

The overall aim is to identify hidden dynamic changes. The original data was given as shown in Table 2. The only data of interest are the data with changes, for example sale amounts of a product, the time stamp, etc, The rest of the static data, such as the weight and the cost of the product can be removed.

Table 2 Original data sets

After cleaning the data, the dynamic attributes are found as shown in Table 3. The dynamic attribute is calculated by finding the difference between sales amounts in one month and sales amounts in the previous month.

Table 3 Dynamic attributes

In the next step, the neutral attributes are pruned. The idea of pruning is to remove redundant dynamic attributes; thus fewer sets of attributes are required when generating rules. The first pruning is based on the single attribute support. In this case, the single attribute support is defined to be 0.5, which means that if an attribute with polarity +, -, or 0 occurs in more than half of total time stamps, it will be pruned. In this case, 429 attributes remain for the rule generation.

In this experiment, dynamic sets are compared based on a simultaneous time stamp. Then the support of sympathetic and antipathetic rules for each dynamic set is calculated. The support is used as the threshold to eliminate unsatisfactory dynamic sets and to obtain the rules from the satisfactory sets.

B. Evaluation and Results

The algorithm was run based on the procedures described in previous sections. Figure 4 shows the plot of sympathetic and antipathetic support. The x –axis represents the support and the y-axis represents the number of rules. This database shows that there are more sympathetic rules than antipathetic rules. The figure shows that increasing support will lead to exponential growth of the rules. As the support reaches 0.05 or 5, as it indicates on the figure, the number of rules is 630. Most of these rules are redundant and have no meaning due to the low support. Figure 5 shows the rule plot with support equal to average value, where the "+support" = "the average of all
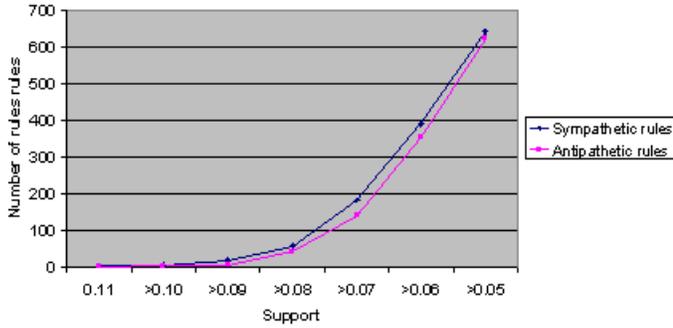
Figure 4 Rule plot with support level = 0

positive records" and "–support" = "the average of all negative records". The number of rules has decreased by applying the support level.
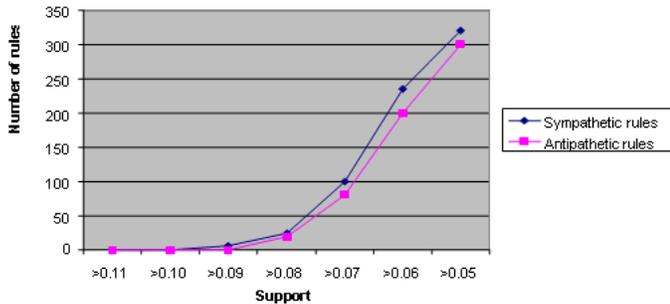

Figure 5 Rule plot with support level = frequent

Table 2 shows the extracted strong rules with support level equal to average value and support larger than 0.08. There are only dynamic pairs so there is no need to do the simulation. The connection can then be put onto concept relational software as
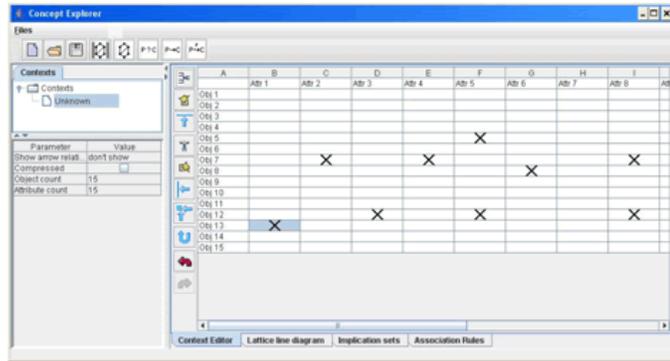

Figure 6. Pictorial scan for ConExp

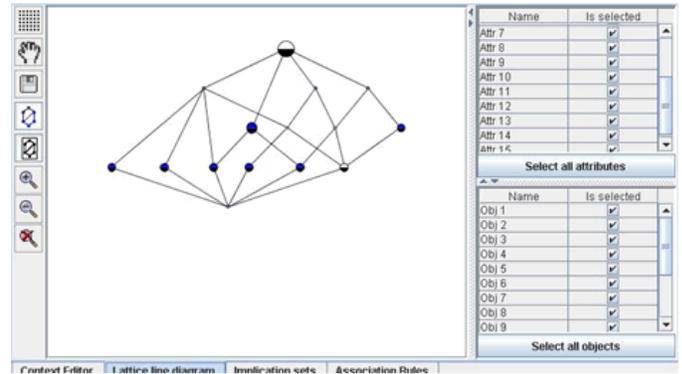ConExp (Conceptual explorer), and can be represented by a lattice as shown in Figure 7.

Figure 7. Lattice generated

Table II
Result generated by the algorithm

Strong rules	Support
Sympathetic	
{C15276179, F030008}	0,093
{J08008008, F060010}	0,089
{A04004004, A05005005}	0,086
{A05005006, C10251104}	0,084
{A04004004, F100020}	0,082
Antipathetic	
{A05005008, C15276179}	0.092
{C10251104, F070010}	0.083
{A05005008, F030008}	0.082

C. Discussion

A priori it is provided some form of causal information, i.e. suggesting a possible direction of causation between two attributes, but there is no basis to conclude that the arrow indicates direct or even indirect causation. The DCM algorithm, on the other hand, shows causality between attributes. Thus, where association rule generation techniques find surface associations, causal inference algorithms identify the structure underlying such associations.

Each type of relationship generated by the *DCM* algorithm provides additional information. The *DCM* algorithm finds four kinds of relationships, each of which deepens the user's understanding of their target system by constructing the possible models. For example, $A_1 \rightarrow^+ A_2$ provides more information than $A_1 \rightarrow A_2$ because the latter indicates that A_1 coexists with A_2. The condition of the rule is not stated (whether sympathetic or antipathetic). A genuine causality such as $A_1 \rightarrow^+ A_2$ provides useful information because it indicates that the relationship from A_1 to A_2 is strictly sympathetic causal.

The rules extracted by DCM can be simulated by using software to model the future behavior. The rules extracted by association algorithm cannot be simulated.

VIII. CONCLUSION

This paper has considered the most fundamental ways to tackle the problems caused by information overload and complexity in retailing. The information system would be available always and everywhere, reacting immediately to any request for guidance or any change in the situation. It would constantly be fed with new information thus cause the

overload. Using semantic web is a popular way dealing with such problems; however the semantic web requires improvement. This paper suggests integration of semantic web and association formal concept analysis method to improve the analysis.

REFERENCES

[1] Abiteboul, S., Buneman, P., and Suciu, D. Data on the Web, Morgan Kaufmann, 2000.

[2] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Inkeri. A. (1996). Fast Discovery of Association Rules, Advances in Knowledge Discovery and Data Mining. The Association for the Advancement of Artificial Intelligence and The MIT Press, 307–328.

[3] Albert, R., Jeong, H. and Barabasi, A.-L., Diameter of the World-Wide Web, Nature, 410, 130-131, 1999.

[4] Andrews, Jonathan; Trites, Gerald, "Net Sales." CA Magazine, Vol.130, No.6:12-15, Aug 1997.

[5] Barabasi, A.-L. and Albert, R., Emergence of scaling in random networks, Science, 286, 509-512, 1999.

[6] Berners-Lee, T., Hendler, J., and Lassila, O. The semantic Web, Scientific American, 29-37, May 2001.

[7] Cheung, D., Han, J., Ng, V. T. & Fu, A.W. (1996).Fast distributed algorithm for mining association rules. In International Conference on Parallel and Distributed Information Systems, Tokyo, Japan, 31–42.

[8] Chimaera (2000). Chimaera Ontology Environment.www.ksl.stanford.edu/software/chimaera

[9] Coyle, R. G. (1996). *System dynamics Modelling: A Practical Approach*, London, Chapman and Hall.

[10] Decker, S.& Melnik, S.. The semantic web: the roles of XML and RDF, IEEE Internet Computing, 4:5, 63-74, 2000.

[11] Decker, S., Mitra, P., and Melnik, S. Framework for the semantic web: an RDF tutorial, IEEE Internet Computing, 4:6, 68-73, 2000.

[12] Ebling, Tom, "The economics of online banking," Target Retailing; Philadelphia; Vol. 24, Issue 2:67-78, Feb. 2001.

[13] Flood, R. (1999). *Rethinking the fifth discipline. Learning within the unknowable*, Ruthledge, London

[14] Forrester, J. W. (1961). Industrial Dynamics: A Major Breakthrough for Decision Makers. *Harvard Business Review*, July-August, 37-66.Fountain, J. E. (2001).

[15] Ginovsky, John, "Bricks can excel at clicks," ABA Bankers News; Washington; Vol. 9, Issue 8, page 8, Apr. 17, 2001.

[16] Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specification. *Knowledge Acquisition* 5: 199-220.

[17] Hackathorn, R.D.Web Farming for the Data Warehouse, Morgan Kaufmann, 2000

[18] Hawn, Matthew, "Stay on the Web", MacWorld, Vol.13, No.4:94-98, Apr. 1996.

[19] Helbing, D., Huberman, B. A., and Maurer, S. M, Optimizing traffic in virtual and real space, in: Traffic and Granular Flow '99: Social, Traffic, and Granular Dynamics, Helbing, D., Herrmann, H. J., Schreckenberg, M., and Wolf, D. E. (Eds.), Springer-Verlag, 2000.

[20] Huberman, B. A. and Adamic, L. A., Growth dynamics of the World-Wide Web, Nature, 410, 131, 1999.

[21] Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E., and Lukose, R. M., Strong regularities in World Wide Web surfing, Science, 280, 96-97, 1997.

[22] Kranzley, Arthur, "Lifestyles fuel Internet banking," Credit Card Management; New York, Vol. 14, Issue 2, page 72, May 2001.

[23] Kosala, R. and Blockeel, H. Web mining research: a survey, ACM SIGKDD Explorations Newsletter, 2, 1-15, 2000.

[24] Higgins, Amy, "Designing with the new Internet," Machine Design; Cleveland; Vol. 73, Issue 14:90-94, July 26, 2001.

[25] Isaac, Peter, "Electronic commerce benefits is frictionless trading," New Zealand Manufacturer, pp.38-39, Feb 1998.

[26] Lane, Andrea, "Success in sight… or site?" Australian Accountant, Vol.66, No.10:22-25 , Nov 1996.

[27] Ling, C.X. and Li, C. Data mining for direct retailing: problems and solutions,

[28] Proceedings of KDD'98, 73-79, 1998.

[29] Liu, J. and Ye, Y. (Eds.) Advances in E-commerce Agents: Brokerage, Negotiation,

[30] Security, and Mobility, Springer-Verlag, 2001.

[31] Liu, J. and Zhong, N. (Eds.) Intelligent Agent Technology: Systems, Methodologies, and Tools, World Scientific, 1999.

[32] Liu, J., Zhong, N., Tang, Y.Y., and Wang, P.S.P. (Eds.) Agent Engineering, World Scientific, 2001.

[33] Long, Johnny, " E-COMMERCE: Doing What's Best for Business," Data Communications, Vol.26, No.16:77-80, Nov 21, 1997.

[34] Malazdrewicz, Michael A., "Navigating on the Net," CA Magazine, Vol.129, No.6:22-26, Aug. 1996.

[35] Marks, Michael, "The Internet: Rewriting the rules of business." Supply House Times, Vol.40, No.11, page 63, Jan 1998.

[36] McGuinness, D.L. and Wright, J. (1998). Conceptual Modeling for Configuration: A Description Logic-based Approach. *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing - special issue on Configuration*.

[37] McKim, Robert, "Dollars and sense on the Web", Target retailing, Vol.20, No.7:30 31, July 1997.

[38] Musen, M.A. (1992). Dimensions of knowledge sharing and reuse.*Computers and Biomedical Research* **25**: 435-467.

[39] Natalya F. Noy and Deborah L. McGuinness Ontology Development 101: A Guide to Creating Your First Ontology.

[40] http://protege.stanford.edu/publications/ontology_development/ontology 101-noy-mcguinness.html

[41] Ontolingua (1997). Ontolingua System Reference Manual. http://www-ksl svc.stanford.edu: 5915 /doc/frame-editor/index.html

[42] Perlow, L. A. (1999) "The Time Famine: Toward a Sociology of Work Time",

[43] *Administrative Science Quarterly,* (44)1, pp. 57-81.

[44] Peterson, R.A., Balasubramanian, S. and Bronnenberg B.J., "Exploring the Implications of the Internet for Consumer Retailing," Journal of the Academy of Retailing Science, Vol. 25, No. 4:329-346, 1997.

[45] Pham, D.T., Wang, Y. & Dimov, S. (2005). Intelligent Manufacturing strategy selection. *Proc 1st Int Virtual Conf on Intelligent Production Machines and Systems* Oxford: Elsevier. 312-318.

[46] Pham, D.T., Wang, Y., & Dimov, S. (2006). Incorporating delay and feedback in intelligent manufacturing strategy selection. *Proc 2nd Int Virtual Conf on Intelligent Production Machines and Systems*. Oxford: Elsevier. 246-252.

[47] Protege (2000). The Protege Project.http://protege.stanford.edu

[48] Rendleman, John, "Customer data means money," Informationweek;, Manhasset; Issue 851:49-50, Aug. 20, 2001.

[49] Richardson, G. P.(1996) System Dynamics. In: Gass, S.I., Harris, C.M. (eds.): Encyclopedia of Operations Research and Management Science. Kluwer Academic Publishers. Boston 656-660.

[50] Richardson, G.P. and Andersen, D.F. Teamwork in group model building. System Dynamics Review, 11 2. (1995), 113-137

[51] Rosenspan, Alan, "The art of the questionnaire," Target Retailing, Philadelphia, Vol.24, Issue 8:42-44, Aug. 2001.

[52] Sandilands, Ben, "The Internet: A tool of the trade?" Australian Accountant, Vol.67, No.11:14-17, Dec 1997.

[53] Senge, P., Kleiner, A., Roberts, C., Ross, R., & Smith, B. (1994). The fifth discipline fieldbook. New York: Doubleday

[54] Simon, H. (1971) "Designing Organizations for an Information-Rich World" in M.

[55] Greenberger (ed.) Computers, Communications, and the Public Interest, Baltimore, MD: The Johns Hopkins Press, pp. 38-52.

[56] Srivastava, J. Cooley, R., Deshpande, M. and Tan, P. Web usage mining: discovery and applications of usage patterns from web data, SIGKDD Explorations, Newsletter of SIGKDD, 1, 12-23, 2000.

[57] Sterman, J. D. (1994). Learning in and about Complex Systems. System dynamics Review. Volume 10 (2-3), 291-330.

[58] Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modelling for a Complex World, Irwin McGraw-Hill, Boston,

[59] Uschold, M. and Gruninger, M. (1996). Ontologies: Principles,Methods and Applications. Knowledge Engineering Review 11(2).

[60] Wolff, K. E. (1994). "A first course in Formal Concept Analysis". F. Faulbaum StatSoft '93: 429–438, Gustav Fischer Verlag.

[61] Yao, Y.Y. and Zhong, N. Mining retail value functions for targeted retailing, Proceedings of the 25th IEEE Computer Society International Computer Software and Applications Conference (COMPSAC 2001), 2001.

[62] Wang, Y. (2007). Integration of data mining with Game theory, Journal of International Federation for Information Processing, Volume 207/2006, Boston: Springer, pp 275-280

[63] Zhong, N., Yao, Y.Y., and Ohsuga, S. Peculiarity oriented multi-database mining, J. Zytkow and Jan Rauch (eds.) Principles of Data Mining and Knowledge Discovery, LNAI 1704, Springer-Verlag, 136-146, 1999.

[64] Zaki, M. J., Parthasarathy, S., Ogihara, M. & Li, W. (1997). New Parallel Algorithms for Fast Discovery of Association Rules, Data Mining and Knowledge Discovery, Vol. 1, No. 4, 343-37.

[65] Zhong, N., Dong, J.Z., and Ohsuga, S. Rule discovery by soft induction techniques,

[66] Neurocomputing, An International Journal, 36:1-4, 171-204, Elsevier, 2000.

[67] Zhong, N. A Study on E-mail Filtering by Uncertainty Sampling and Relation Learning, Technical Report, Yamaguchi University, 2000.

[68] Zhong, N., Yao, Y.Y., and Kakemoto, Y. Automatic construction of ontology from text databases, N. Ebecken and C.A. Brebbia (Eds.) Data Mining, 2, WIT Press, 173-180, 2000.(b)

[69] Zhong, N. Knowledge discovery and data mining, The Encyclopedia of Microcomputers, 27, Supplement 6,235-285, Marcel Dekker, 2001.

The databases were contributed by many researchers, mostly from the field of machine learning and data mining and collected by the machine learning group at the University of California, Irvine. Two of the datasets, MCslom and ASW, are taken from real life. These data sets are described briefly below

Cystine Database: This data arises from a large study to examine EEG correlates of genetic predisposition to alcoholism. It contains measurements from 64 electrodes placed on the scalp sampled at 256 Hz (3.9-msec epoch) for 1 second.

Weka base: This dataset contains time series sensor readings of the Pioneer-1 mobile robot. The data is time series, multivariate. The few are binary coded 0.0 and 1.0. Two categorical variables are included to delineate the trials within the datasets. The data is broken into "experiences", in which the robot takes action for some period of time and experiences a controlled interaction with its environment.

Market basket: Classical association data mining, used in WeKa analysis. It consists of 100 different transactions.

Bank data: Includes 600 instances of bank transactions and 12 attributes in each instance.

ASW: This data consists of real life data. This dataset contains 65536 attributes of metal manufacturing, with 8 records in each attribute.

Mclosom: Manufacturing database for logistics. 72 time stamps and 50 attributes for 5 different classes.