Data Integration Potentiometer in DERMIS

Eli Rohn
Department of Information Systems, NJIT
elirohn@njit.edu

ABSTRACT

Dynamic Emergency Response Management Information Systems must integrate data from heterogeneous and autonomous resources. We propose a mathematically based approach for evaluating and quantifying the potential of a successful automatic integration between pairs of data resources. The integration potential, expressed as a percent, should influence assumptions and procedures of emergency response planners.

Keywords

Data Integration, Payoff Function, Emergency Response

INTRODUCTION

A response to an emergency more often than not involves several organizations that under normal operations are loosely connected or entirely unrelated. For example – responding to the 1989 Exxon Valdez spill required the cooperation of private enterprises and government agencies at the local and Federal levels (Harrald, Marcus, & Wallace, 1990). Responding to the 1993 terrorist attack on the World Trade Center (WTC) required the cooperation of several government agencies such as the New York Police Department (NYPD) the New York Fire Department (NYFD), the Port Authority, and private enterprises such as the WTC building's management, ambulances and hospitals (The 9-11 Commission, 2004).

Adequate management of the response to a calamity requires good coordination of efforts and resources. Dynamic Emergency Response Management Information Systems (DERMIS) (Turoff, Chumer, Van de Walle, & Yao, 2004) have the potential in aiding and optimizing the management process. One of the requirements from such a system is the ability to integrate data from responding organizations and structure it in a manner that yields usable information for the management team. The DERMIS paper introduces eight design principles. Two of them – principle 5 and principle 6 rely on adequate and timely data integration. These design principles reads "Up-to-Date Information and Data: Data that reaches a user and/or his/her interface device must be updated whenever it is viewed on the screen or presented verbally to the user… Link Relevant Information and Data: An item of data and its semantic links to other data are treated as one unit of information that is simultaneously created or updated."(Turoff, Chumer, Van de Walle, & Yao, 2004).

Automatic integration of heterogeneous data sources is a highly desirable feature in DERMIS. The Related Work section provides an overview of approaches and technologies aimed at accomplishing this goal. The difficulties outlined here necessitate thorough consideration when designing a DERMIS and when using it operationally. The review shows that to date there is no software solution that ensures accurate automatic integration of heterogeneous data sources. This jeopardizes the functionality of DERMIS designed with automatic integration in mind.

We propose an approach for evaluating and quantifying the potential success of automatic integration between pairs of data resources in the event the two need to be integrated into DERMIS. An early evaluation will indicate if automatic integration is possible and to what degree. Poor automatic integration potential means that during the planning phase emergency response organizations should allocate adequate technical personnel and domain experts to support manual mapping between data schemas, or assume that data integration will not be possible at the outset of the response and amend their assumptions and procedures accordingly.

RELATED WORK

Integration of computerized information systems spans over four decades. Gosden summarized some of the fundamental requirements of the problem and characteristics of the solutions: "There is a need for a Data Definition Language (DDL) to describe files in both syntactic and semantic aspects in such a manner as to provide for a suitable mapping to appropriate physical structure" (Gosden, 1969). The federated architecture for database systems proposed by (McLeod & Heimbigner, 1980) was supposed to address some of the shortcomings of completely centralized databases by proposing an architecture that supports the physical and logical decentralization of databases. (Motro & Buneman, 1981) introduced a fundamental conceptual procedure for integrating component views into a 'superview', later known as Global Schema. The procedure provides a unified view of two or more databases while preserving their physical independence. The Computer Corporation of America first built a system for integrating preexisting, heterogeneous, distributed databases as part of the Multibase project in the early 80's (Doan, Madhavan, Domingos, & Halevy, 2002; Smith et al., 1981). The common thread among these approaches is the complexity of data integration. Their approaches and subsequent inability to produce a complete and accurate integrated view reflects the common constraining forces that continue to make automatic integration an active area for research.

The proliferation of networked information systems and specifically the Internet created a need for a universal data transfer language merging document-centric Web pages with a data-driven infrastructure. The concepts of self-describing semi-structured data took center stage, giving rise in 1996 to the eXtensible Markup Language (XML) due to its fit to self-describing data representation (Cseri, Layman, Lovett, Paoli, & Schach, 1997). XML is "self-describing, so that structured data expressed in XML may be manipulated by software that doesn't have previous knowledge of the underlying meaning behind the data. XML provides a file format for representing data and can be extended to contain a description of its own structure" (Cseri, Layman, Lovett, Paoli, & Schach, 1997). XML yielded a barrage of data integration projects and a flood of integration related publications. For example – The Distributed Information Search Component (DISCO) (Tomasic, Amouroux, & Bonnet, 1997); Lorel (Abiteboul, 1997); STRUDEL (Mary Fernandez, Florescu, Kang, Levy, & Suciu, 1997; Mary Fernandez, Florescu, Levy, & Suicu, 2000); YAT (Sophie Cluet, Delobel, Siméon, & Smaga, 2001; Sophie Cluet & Siméon, 1999); Xyleme (Aguilera, Cluet, Veltri, Vodislav, & Wattez, 2001); The number of commercial applications that claim integration capabilities, especially XML processing capabilities, is reaching a record high (Bourret, 2003). XML has lowered the barrier to entry for data exchange, but with it came a flood of proposed standards. A broad range of professional organizations are promoting and developing XML-based standards. Gartner identified well over 2,000 proposed XML standards (Knox, 2002).

XML did not provide a panacea for automatic integration of heterogeneous and autonomous data sources. A fundamental issue is incompatible semantics, which requires disambiguation. XML data structures exhibit numerous types of ambiguities: Word Sense ambiguity; Structural ambiguity; Projection ambiguity; Referential ambiguity; Resolution ambiguity (Rohn & Klashner, 2004).

Ongoing integration research focuses on the usage of ontologies for disambiguation. Ontologies aim at capturing static domain knowledge in a generic way and provide a commonly agreed upon understanding of that domain, which may be reused and shared across applications and groups (Chandrasekaran, Josepheson, & Benjamins, 1996). Ontologies do not provided a working solution for automatic integration. For example, the DARPA Agent Markup Language (DAML) Program officially began in August 2000 and was terminated in 2003 due to lack of progress. Resource Description Framework (RDF), a graph-based model for describing Internet resources (Lassila & Swick, 1999; W3C, 2004), doesn't resolve ambiguities of data structures. The Web Ontology Language (OWL) (Bechhofer *et al.*, 2004; W3C, 2004) supports ontologies, but the number of ontologies keeps on increasing and presents a problem similar to the one created by the uncontrolled development and proliferation of XML based "standards". Consequently there is a need to map local ontologies to a global ontology (Calvanese, De Giacomo, & Lenzerini, 2001) for an ontology-based integration to work. Such a mapping presents challenges that are at least as complex as mapping among heterogeneous data structures. To date there is no software solution that ensures accurate automatic integration of heterogeneous data sources. This jeopardizes the functionality of DERMIS designed with automatic integration in mind. There exists a need to evaluate existing and emerging technologies and vocabularies for their integration potential. The following sections explain our approach.

MEASURING INTEGRATION POTENTIAL

Semantic Heterogeneity is the term used by Halevy to describe the differences in same domain database schemas developed by independent parties. The schemas will almost always be quite different from each other (Halevy et al., 2005). We estimate the magnitude of semantic heterogeneity expressed in percentage using mathematical analysis.

Signal-meaning mappings in the form of lexical matrixes are central to our approach. Lexical matrixes associate words and meanings (Hurford, 1987). Lexical matrixes are not bound to any domain or language. They can be used for signal-meaning mappings between a dog and its handler, jungle drummers and tribesmen, or signal-meaning mapping of humans communicating using natural language. "A lexical matrix is a convenient description of arbitrary relations between discrete forms and discrete concepts. [It] is an integral part of the human language system. It provides the link between word form and word meaning. A simple lexical matrix is also at the center of any animal communication system, where it defines the associations between form and meaning of animal signals." (Komarova & Nowak, 2001). Lexical matrixes have been used with languages other than English in the Information Retrieval field (Chklovski, Mihalcea, Pedersen, & Purandare, 2004).

Let us organize the vocabulary of a public ("exposed") data source by defining a *Source matrix.* Let us arrange the vocabulary of a Global Schema in a *Global Schema matrix.* The matrixes are different due to their semantic heterogeneity. Every matrix contains a number of arbitrary data elements (fields, XML tags, etc.) used to communicate information about a number of objects and concepts in their common domain. Each data source / global schema has an active matrix *Q* and a passive matrix *P*. The active matrix is used to broadcast ("expose" / make publicly available) data elements. The passive matrix is used to make sense of incoming data elements such as when integrating external data structures into a global schema or when a local schema is queries

Q and P are stochastic matrixes whose entries are [0,1] and each of their rows sum to one. Two matrixes define the language of a data source as L=(Q , P) whereas the Global Schema's language L' = (Q' , P').

A *payoff* (Nowak & Krakauer, 1999) is the number of objects that can be communicated between the data source and the global schema weighted by their probability of correct communications.

Let Q_{ij} denote the probability that a data source will refer to object *i* by using element *j*. Let P_{ji} denote the probability that an integration system with a Global Schema will interpret data element *j* as object *i*. Let *n* be the number of objects and let *m* be the number of data element.

$$Payoff(L , L') := ½ \Sigma_{(I =1 \text{ to } n)} \Sigma_{(J =1 \text{ to } m)} \quad Q_{ij} * P'_{ji} + Q'_{ij} * P_{ji} = \qquad (1.1)$$

$$Payoff(L , L') := ½((PQ') + (P'Q)) \qquad (1.2)$$

$$\text{Max}(Payoff) = \text{Min}(n , m) \qquad (1.3)$$

We can see in (1.3) that the maximum payoff is the smaller of the two (n , m). The payoff absolute maximum value can be reached only when *n* = *m*.

Payoff Demonstration

	Active	Passive
RETS	Q	P
MISMO	Q'	P'

Table 1: The 4 Matrixes

We use empirical representative data structures obtained from real-world samples to demonstrate the payoff. We also use WordNet (Miller, 1995; WORDNET, 2005), a lexical database for the English language, to determine the number of meanings of some word in our matrixes. For the purpose of this article we initially chose two emergency related schemas: The OASIS Emergency CAP (OASIS, 2005) and the DOT HAZMAT incident report (DOT, 2005). Although tempting at first sight, we opted not use them after all. We observed that schemas from the emergency management domain have little overlap compared to schemas form other domains, such as the Real Estate domain, making the emergency schemas susceptible to low levels of agreement from the outset. This is especially true for the two aforementioned emergency schemas. The CAP schema has 43 data elements composed of 48 unique words. The DOT schema has over 1700 data elements composed of different combinations of 184 unique words. These differences indicate there is significant semantic heterogeneity between the two schemas. Such opening conditions make it apparent that the payoff is extremely low, bordering zero. We decided to use data from the Real Estate domain (Rohn & Klashner, 2004) due to their known similarities, as our objective here is to demonstrate the utility of the payoff function. Our first data source is the Real Estate Transaction Standard (RETS), expressed in XML (RETS, 2004); the other is Property Information Data Dictionary, sponsored by the Mortgage Industry Standards Maintenance Organization's (MISMO), a de-facto standard in the Real Estate industry, expressed as a glossary of terms and engineered by MISMO into XML (MISMO, 2001, 2005). The RETS data structure uses 211 unique words. MISMO has 105 unique words. The two schemas have six words in common. Two of them are "Type" and "ID", and they participate in the lexical matrixes we use to demonstrate our approach.

Data integration is a two-way process. First – the integration from the source to the global schema; second the querying from the global schema to the source. Table 2 and Table 3 address the first process – from the source

(RETS) to the Global Schema (MISMO). Table 4 and Table 5 address the second process – from the global schema to the source.

		OBJECTS					
		ID	**Address**	**Name**	**File**	**Type**	**∑ Row**
RETS	Type	0.2511	0.1322	0.1515	0.3402	0.1250	1.0000
	RE	0.2266	0.3280	0.1413	0.1459	0.1581	1.0000
	Code	0.1424	0.2098	0.2574	0.2244	0.1660	1.0000
	ID	0.3333	0.2183	0.2760	0.0498	0.1225	1.0000
	Present	0.1765	0.1070	0.2063	0.2201	0.2902	1.0000
	Currency	0.2750	0.2693	0.0432	0.2478	0.1647	1.0000
	Office	0.2366	0.1700	0.1977	0.2408	0.1549	1.0000

Table 2: RETS Active Matrix Q

Table 2 lists vertically the seven most frequently used words (XML tags or parts thereof) in RETS. The table lists horizontally the five most frequently used concepts in MISMO. Each cell Q_{ij} represents the probability that RESTS refer to object *i* by using word *j* The probabilities were assigned randomly, except for the cells Q5,1 and Q1,4. The first one uses the RETS signifier "type". The word "type" has 8 meanings in WordNet, thus we assign the probability 1/8 = 0.1250. "ID" has three meanings in WordNet, so the probability we assigned to it is 1/3 or 0.3333. We recognize that the semantic importance of a sense of a word should influence it probability assignment. Yet we opted for a simplified assignment of probabilities for two reasons: first – to keep things simple at this stage. Second, unlike natural language text corpora, schemas are comprised mostly of nouns, lacking verbs and adjectives that help assign sense importance in natural language processing. For example, the DOT IncidentReport has 11 words out of 184 that could qualify as verbs. RETS has 28 words out of 211 that could qualify as verbs. In either examples the potential verbs are usually used as nouns, such as with the words “date” and “last”. Additionally, schema elements do not form complete sentences, as one finds in text corpora. We don’t know of exiting techniques for weighted probability assignments that will work as intended under such skewed circumstances.

		MISMO Words					
		ID	**Address**	**Name**	**File**	**Type**	**∑ Row**
RETS	Type	0.1445	0.2933	0.1700	0.2672	0.1250	1.0000
	RE	0.2061	0.1832	0.2818	0.1303	0.1986	1.0000
	Code	0.1559	0.1524	0.1351	0.3174	0.2392	1.0000
	ID	0.3333	0.1027	0.2407	0.0361	0.2872	1.0000
	Present	0.2591	0.1489	0.1664	0.2250	0.2006	1.0000
	Currency	0.1200	0.2354	0.2450	0.1328	0.2668	1.0000
	Office	0.3122	0.1849	0.1353	0.1250	0.2425	1.0000

Table 3: MISMO Passive Matrix P’

Table 3 lists the same concepts shown in table 2. The difference is that each cell Q_{ij} represents the probability that MISMO interprets word *j* (used by RETS) as object *i*. That is, P'_{ji} denote the probability that an integration system with a MISMO based global schema will interpret the RETS element *j* as object *i*. The probabilities were assigned randomly, except for the cells Q5,1 and Q1,4 as explained earlier.

		MISMO					
		ID	**Address**	**Name**	**File**	**Type**	**∑ Row**
RETS	Type	0.1923	0.2094	0.1570	0.3162	0.1250	1.0000
	RE	0.3071	0.1665	0.2348	0.0749	0.2167	1.0000
	Code	0.1163	0.3443	0.1153	0.2345	0.1895	1.0000
	ID	0.3333	0.1791	0.0010	0.1974	0.2892	1.0000
	Present	0.3446	0.1213	0.1581	0.1189	0.2571	1.0000
	Currency	0.1386	0.1900	0.3090	0.1169	0.2456	1.0000
	Office	0.2613	0.2079	0.1888	0.0806	0.2614	1.0000

Table 4: MISMO Active Matrix Q'

		OBJECTS					
		ID	**Address**	**Name**	**File**	**Type**	**∑ Row**
RETS	Type	0.2345	0.3259	0.1139	0.2007	0.1250	1.0000
	RE	0.3506	0.2585	0.1618	0.1179	0.1112	1.0000
	Code	0.1981	0.1656	0.1837	0.2271	0.2253	1.0000
	ID	0.3333	0.1112	0.2478	0.1814	0.1263	1.0000
	Present	0.1835	0.2196	0.2490	0.2322	0.1156	1.0000
	Currency	0.2016	0.1635	0.1964	0.1384	0.3000	1.0000
	Office	0.2950	0.2895	0.2068	0.1122	0.0964	1.0000

Table 5: RETS Passive Matrix P

Table 4 denotes the probabilities that MISMO will use tag *i* to denote object *j*. Table 5 denotes the probability that RETS will interpret MISMO's object *j* as RETS' tag *i*.

Now that we have all four matrixes we can calculate the payoff(L, L') using formula (1.1). Table 5 illustrates the calculations steps. Maximum payoff is determined by formula (1.3) as min(7,5) = 5. The calculated payoff is about 29%.

I	J	Qij	P'ji	Q'ij	Pji	Qij*P'ji	Q'ij*Pji	Total
1	1	0.2511	0.1445	0.1923	0.2345	0.0363	0.0451	0.0814
1	2	0.1322	0.2933	0.2094	0.3259	0.0388	0.0682	0.1070
…	…							
7	5	0.1549	0.2425	0.2614	0.0964	0.0376	0.0252	0.0628
						Total payoff		1.4339
						Maximum payoff		5.00
						% Payoff		**28.6800**

TABLE 5: Payoff Calculation (truncated)

We created similar matrixes of for RETS and MISMO having the size 30x30. We executed over a dozen Monte-Carlo experiments in order to obtain a payoff average. The average payoff is 3.90%. The all-inclusive payoff matrix of RETS x MISMO (of size 211 x 105) will yield a much smaller payoff due to the size of their actual semantic heterogeneity.

DISCUSSION

The payoff function reflects the total amount of information that L can convey to L' and vice versa. We used two highly similar representative data sources from the same domain assuming it will create a large payoff due to an anticipated significant number of common concepts. An early indication of a possible problem was the 100% difference in vocabulary size between the two data models. RETS has 211 words while MISMO uses only 105. An additional early warning sign was the fact that RETS and MISMO have only six words in common, among their first 30 most used words. An artificially small payoff matrix (7x5) yields a dismal payoff of about 28%, indicating the presence of significant semantic heterogeneity (72%). A larger matrix (30x30) yields 4% payoff, which can be interpreted as having 96% semantic heterogeneity. Either payoff is too small for meaningful automatic integration. The results indicate that mapping between these two sources do not preserve the meaning of common concepts. Meaning preservation is required for meaningful integration.

By extrapolation we conclude that integration between two distinct domains, such as heavy equipment (equipment properties, availability, location etc.) and electric grid (location of available power, amount of available power, location of high-voltage live down wire, etc.) will yield a payoff that is not better than the one we examined. This should be of great concern to emergency response management teams who plan on relying on DERMIS to help manage the response. With the current technology there seems to be one practical solution: the creation of an emergency response data vocabulary standard. Standards work very well in other domains, such as in manufacturing (e.g., EDI) and in international banking (e.g., SWIFT). The development of such a standard requires additional research, as well as obtaining the consensus of multiple organizations, private, public and government agencies. Planners of specialized standards that should map to a global emergency response standard should keep in mind that a payoff's absolute maximum value can be reached only when $n = m$. It unrealistic to think it can happen often by chance. Data models constructed upon proposed standards should be designed with the goal of keeping n and m as close as possible. Standards have limitations, and one of them is rigidity, so they might not be the perfect solution.

We believe that there exists an additional solution, but it is entirely theoretical at this point. It revolves around the creation of an iconic language that uses a small number of elementary objects and a set of rules that allow the combination of such elementary objects into meaningful complex objects. Meaningful to a computer, that is. This approach abolishes the dependence on natural language thus potentially eliminating the problem of ambiguity, while allowing for ad-hoc creation of meaningful unambiguous complex objects when need arises. A tiny iconic language we created in a separate ongoing unpublished research yields payoff that is greater than 70%. However, further extensive research is required in order to realistically evaluate its payoff potential and its implementation feasibility.

CONCLUSION

DERMIS must integrate data from heterogeneous and autonomous resources ad-hoc. Fully automated integration is the preferred method. We used a representative sample of publicly available data structures to evaluate their automatic integration potential using lexical matrixes and a payoff function. Even data structures from the same domain yield small payoffs, indicating poor suitability for automatic integration. Managers of emergency response teams must take these current limitations into consideration when designing DERMIS and planning for required support resources.

ACKNOWLEDGMENTS

I thank Dr. Robert Klashner of NJIT's IS Department for his editorial help in the work related section and Dr. Denis Blackmore of NJIT's Math department for numerous informal discussions we had on the subject.

REFERENCES

1. Abiteboul, S. (1997). *Querying semi-structured data.* Paper presented at the 6th International Conference on Database Theory (ICDT'97).
2. Aguilera, V., Cluet, S., Veltri, P., Vodislav, D., & Wattez, F. (2001). Querying XML Documents in Xyleme. Retrieved July 7, 2005, from http://www.haifa.il.ibm.com/sigir00-xml/final-papers/xyleme/XylemeQuery/XylemeQuery.html
3. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., et al. (2004, 10 February 2004). OWL Web Ontology Language Reference. *W3C Recommendations* Retrieved July 7, 2005, from http://www.w3.org/TR/owl-ref/
4. Bourret, R. (2003, January 1, 2005). XML Database Products. Retrieved 05/01/2005, from http://www.rpbourret.com/xml/XMLDatabaseProds.htm
5. Calvanese, D., De Giacomo, G., & Lenzerini, M. (2001). *A Framework for Ontology Integration.* Paper presented at the International Semantic Web Working Symposium.

6. Chandrasekaran, B., Josepheson, J., & Benjamins, V. R. (1996). Ontologies: What are they? Why do we need them. *IEEE Intelligent Systems, 14*(1), 20-26.
7. Chklovski, T., Mihalcea, R., Pedersen, T., & Purandare, A. (2004). *The Senseval-3 Multilingual English-Hindi lexical sample task.* Paper presented at the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Barcelona, Spain.
8. Cluet, S., Delobel, C., Siméon, J., & Smaga, K. (2001). *Your Mediators Need Data Conversion.* Paper presented at the The 27th VLDB Conference, Rome, Italy.
9. Cluet, S., & Siméon, J. (1999). *Using YAT to Build a Web Server*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
10. Cseri, I., Layman, A., Lovett, C., Paoli, J., & Schach, D. (1997). Building XML Parsers for Microsoft's IE4. Retrieved January, 2006, from http://www.xml.com/pub/a/w3j/s3.paoli.html
11. Doan, A., Madhavan, J., Domingos, P., & Halevy, A. (2002). *Learning to Map between Ontologies on the Semantic Web.* Paper presented at the The 11th International WWW Conference, Hawaii.
12. DOT. (2005). DOT Form 5800.1 Hazardous Materials Incident Report XML Schema. Retrieved 21 January, 2006, from http://hazmat.dot.gov/pubs/pubs.htm
13. http://hazmat.dot.gov/regs/rules/final/69fr/69fr-67384.htm
14. Fernandez, M., Florescu, D., Kang, J., Levy, A., & Suciu, D. (1997). *STRUDEL: a Web site management system*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
15. Fernandez, M., Florescu, D., Levy, A., & Suicu, D. (2000). Declarative Specification of Web Sites with STRUDEL. *Very Large Data Bases (VLDB), 9*(1), 38-55.
16. Gosden, J. A. (1969). Report to X3 on data definition languages. *ACM SIGMOD Record, 1*(2), 14 - 25.
17. Halevy, A. Y., Ashish, N., Bitton, D., Carey, M., Draper, D., Pollock, J., et al. (2005). *Enterprise information integration: successes, challenges and controversies*. Paper presented at the Conference Name|. Retrieved Access Date|. from URL|.
18. Harrald, J., Marcus, H., & Wallace, W. (1990). Exxon Valdez: An Assessment Of Crisis Prevention And Management Systems. *INTERFACES, 20*(5), 14-30.
19. Hurford, J. (1987). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua, 77*(2), 187-222.
20. Knox, R. E. (2002). *Here's What's Wrong With XML-Defined Standards* (Gartner Research Note No. Number: COM-18-7987): Gartner Group.
21. Komarova, N., & Nowak, M. A. (2001). The Evolutionary Dynamics of the Lexical Matrix. *Bulletin of Mathematical Biology, 63*, 451-484.
22. Lassila, O., & Swick, R. R. (1999). Resource Description Framework (RDF) Model and Syntax Specification [Electronic Version]. *W3C Recommendation, 1999*. Retrieved July 2002 from http://www.w3.org/TR/REC-rdf-syntax/.
23. McLeod, D., & Heimbigner, D. (1980, 1980). *A federated architecture for database system.* Paper presented at the National Computer Conference, Anaheim, CA.
24. Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39-41.
25. MISMO. (2001, 21 Feb 2001). Press Release. Retrieved April 2004, from http://xml.coverpages.org/ni2001-02-27-c.html
26. MISMO. (2005, January 24 - 27, 2005). MISMO Technology Workgroup Meetings. Retrieved 02 May 2005, 2005, from http://webster.mismo.org/mismo/mismo_meeting_report_0105.cfm
27. Motro, A., & Buneman, P. (1981, April 1981). *Constructing superviews.* Paper presented at the SIGMOD International Conference on Management of Data, Ann Arbor, Michigan.
28. Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences of the United States of America, 96*(14).
29. OASIS. (2005, September 08, 2005). OASIS Advances CAP and Emergency Data Exchange Language (EDXL) Specifications. Retrieved 21 January, 2006, from http://xml.coverpages.org/ni2005-09-08-a.html
30. RETS. (2004). RETS XML Proposed Standard. Retrieved April 2004, 2004, from http://www.rets.org/

31. Rohn, E., & Klashner, R. (2004, August 2004). *Hidden Disorder in XML Tags.* Paper presented at the Proceedings of the Americas Conference on Information Systems, New York, NY.
32. Smith, J. M., Bernstein, P. A., DayaI, U., Goodman, N., Landers, T., Lin, T., et al. (1981). *Multibase - integrating heterogeneous distributed database systems.* Paper presented at the National Computer Conference.
33. The 9-11 Commission. (2004). Heroism and Horror. In *The 9-11 Commission Report*: U.S. Government Printing Office.
34. Tomasic, A., Amouroux, R., & Bonnet, P. (1997). *The distributed information search component (Disco) and the World Wide Web.* Paper presented at the Proceedings of the 1997 ACM SIGMOD.
35. Turoff, M., Chumer, M., Van de Walle, M., & Yao, X. (2004). The Design of a Dynamic Emergency Response Management Information System (DERMIS). *The Journal of Information Technology Theory and Application (JITTA), 5*(4).
36. W3C. (2004). OWL Web Ontology Language Overview. Retrieved 02 May 2005, 2005, from http://www.w3.org/TR/owl-features/
37. WORDNET. (2005). WordNet Website. 2005, from http://www.cogsci.princeton.edu/cgi-bin/webwn1.7.1