Using Context in Semantic Data Integration

George Karabatis
Department of Information Systems
University of Maryland, Baltimore County (UMBC)
Baltimore, MD 21250, USA
georgek@umbc.edu

Abstract: The lack of semantic information on disparate data sources in today's heterogeneous information systems poses a significant problem in formulating a multi-system query, which accurately reflects user intentions. This paper presents a metadata approach that exploits context information and semantic network technology to perform an intelligent semantic integration of diverse data. Specifically, it enhances, augments, and refines the original user query to convey semantically relevant context in an accurate manner. A prototype system expressing this idea is described, and preliminary experiments using environmental data are presented.

Introduction

Corporations, governments, researchers, and the general public increasingly deal with vast amounts of stored data, in search of answers to their questions. Since there is no single universal repository that contains everything (at least for now), the quest for information has been directed towards multiple repositories that store different types of data which are used to answer a user query. However, the disparities of diverse data repositories constitute a hard problem for the interoperability of heterogeneous information systems. A major problem is to identify semantically related objects in different repositories, target the search to the most appropriate data sources, and create an integrated view of the participating data sources [KS96]. Many times users are not even aware of the existence of repositories that may contain data quite relevant to their query, and unfortunately these data sources are not incorporated in the submitted queries. Therefore, they unknowingly deprive themselves from accessing sources which may provide not only additional but more precise and succinct results to their search. We propose a solution to this problem by exploiting context information and semantic network technology to recommend to the users additional data sources that enhance their original query and thus, improve the quality of the result.

This paper focuses on the influence of context and semantic networks to the problem of semantic data integration. It is important to understand that when users access multiple repositories in search of information, they tend not to be explicitly aware of their context - instead they take it for granted. Contextual information plays a significant role in the transformation of the user query into a new one, which identifies, captures, and presents user intentions in a better form, because it incorporates additional information reflecting the context.

Semantic networks have been used to help in semantic data integration: They are graphs containing nodes (corresponding to data sources) and edges (identifying relationships between data sources). Thus, semantic networks express the degree of semantic relevance between related data sources. They are quite useful in revealing additional data sources semantically related to the ones in the original user query that the user might not be aware of. We claim that the combination of context information with semantic networks constitutes a powerful approach to the problem of semantic data integration. The contributions of this paper are:

- Exploitation of various types of context to enhance semantic knowledge on background information of user queries.
- Utilization of semantic networks to recommend additional data sources based on context.
- Transformation of the original user query into a new one, which conveys information from appropriate contexts and semantic networks, reflects user intent accurately and retrieves more precise results.

We have implemented a prototype system to use as a test-bed and proof-of-concept of our solutions, where an initial user query is augmented and transformed to return additional and/or more precise information. The increased amount of information has a wider range (since it spans over additional, yet semantically related data sources that the user may not be aware of - automatically increasing its scope) and at the same time is more precise (since it takes various types of context into consideration), resulting in more integral and succint information retrieval from multiple heterogeneous data sources. We have also performed a number of experiments which verify our methodology.

The next section presents background information on context and semantic networks. Following related work in the areas of context and semantic data integration, we present our methodology on merging context with semantic network technology and transforming a user query. An evaluation of the prototype implementation is discussed and finally we outline our conclusions and further work.

Background on Context and Semantic Networks

This section provides a brief and necessary background information on context and semantic networks. It lays the foundation on which we build our contributions for semantic data integration based on context and semantic networks.

On Context

An isolated statement conveys a certain meaning to the reader. However, if it is considered together with other nearby statements, its meaning is “in context” and the statement becomes more clear, precise, and specific. The context qualifies the statement and makes it more accurate. On the other hand, if it is extracted from the nearby statements it may convey an imprecise meaning, sometimes a distorted meaning, completely different from the “in context” meaning. It is now considered to be “out of context”. We carry this analogy to the user query and we describe a methodology to put the user query “in context” by exploiting context information to enrich and augment the user query.

In general, we know what we mean when we talk about context in any situation, but it is not easy to actually define context precisely, because context is more subjective rather than objective. Different people mean different things when they refer to context. As such, different people define different contexts. But can we come up with a universal definition of context? There has been significant amount of research work based on, and using context, and multiple definitions of context. Within the scope of this paper, we define context as *everything that surrounds us, the environment that contains background information in an interaction between a user and a computer*. It contains background information about the user, the computer system, the query that is submitted, that is, all environment aspects that exist and surround such interaction. It is the state of the user, along with the physical and computational environment surrounding the user. The history of user actions is also considered part of context.

When a user poses a query to an information system consisting of multiple heterogeneous information repositories, the more information we can provide to the system about the user, the query, and the semantics of the user's quest, then the better result set we can expect to receive from the system. Unfortunately, most systems do not consider such information and rely only on user input and a description of information stored in participating repositories; they may at most consider information stored in ontologies and/or apply semantic techniques. However, these approaches do not consider the background or the environment where such query is posed. This gap can be filled when context information is provided to the system along with the user query. Context qualifies each user request with additional information about the user and the environment, and supplies the system with supplementatry, yet important input towards a more precise answer.

On Semantic Networks

A semantic network consists of nodes corresponding to data sources and edges corresponding to relationships between sources. They have been used in philosophy, psychology and more recently in information and computer science. Sowa gives a descriptive outline on the types and use of semantic networks in different disciplines in [S06, S92]. Semantic networks have long been used to represent relationships [M61]. They are also known as associative networks, Bayesian networks, or causal networks. Semantic networks can be used to recommend additional data sources of possible interest to users. Information on applying semantic network techniques for the quest of relevant but undiscovered information in the environmental domain can be found in [CGK07].

The semantic networks used in this paper are directional graphs that contain nodes, representing data sources, and edges with a relevance score, identifying the degree of relevance between the two connected nodes. Figure 1 illustrates a semantic network identifying data sources related to fish species. If a user is interested in the 'Fish Information' data source, the semantic network will investigate Stream Temperature, Impervious Surfaces, and Pollution data sources, and possibly recommend some of them to the user.

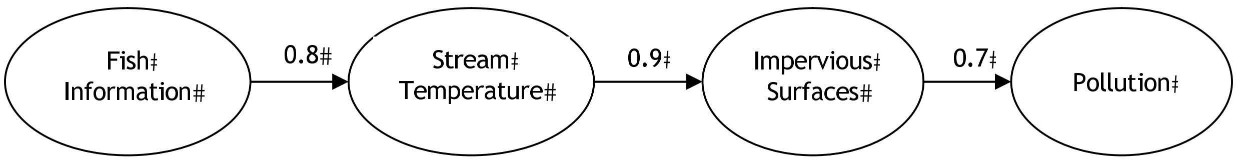

Figure 1. An example semantic network

The relevance score is used to infer the relevance between any two nodes in the network. It identifies the probability of a user interested in data on the source node to access data on the target node. Each relevance score is a conditional probability independent of each other [R94]. For example, the relevance score between Fish Information and Stream Temperature is the conditional probability of a user interested in stream temperature given that he or she is interested in fish. Pearl has used probabilities in semantic networks and has applied statistics and probability in causal semantic networks - also called causal diagrams or belief networks [P00, P88]. Pearl used various techniques and statistical methods to derive a causal or belief network from observed data.

There can be more detailed types of semantic relationships (cause-effect, is-a, and is-part-of) [FKN91, SAK03] linking two nodes. Also closeness, or proximity, defined as similarity between data sources can also be used to identify relationships [SAK03]. The edge can also denote multiple relationships between two nodes, separated by e.g., different color. For example, red edges may represent is-a relationships while blue edges may represent cause-effect relationships. The strength of the relationship is visually represented as the thickness of the edge in [ZRG05], in yet another version of semantic networks called relevance networks. In this paper, we use probabilities as the only information on the edges. Additional meanings and semantics on edges are beyond the scope of this paper.

Users in search of data sources may perform a keyword search or submit a regular SQL query, which in turn will find data sources that satisfy the submitted conditions. These data sources are called exact answers. Semantic networks enhance and augment the exact answers with additional sources, semantically relevant to the exact answers, which the user might not be aware of. This can be achieved by supplying all data sources whose relevance score (conditional probability) with the exact answers is higher than a user-defined threshold.

Related Work

A significant part of scientific literature is related to the use of context not only in information related disciplines but also in the social sciences such as psychology and sociology. Pomerol and Brezillon examine context and identify it as external, contextual, and proceduralized [PB99]. Bazire and Brezillon made an analysis on 150 definitions of context found on the web, in different domains of cognitive science, and concluded that the definition of context depends on the field of knowledge it belongs to [BB05]. For a comprehensive examination of context in artificial intelligence, databases, communication, and machine learning, see [B96].

Lieberman and Selker present context in computer systems and describe it as "everything that affects the computation except the explicit input and output" [LS00]. This is a definition in line with what we also advocate. Beyond the conceptual definition of context, there is enough research performed within the scope of data integration and interoperability using context. Goh et al. present a mediator approach where semantic conflicts are detected and reconciled by a context mediator [GBM99]. Context has also been used as an aid in defining semantic similarities between database objects [KS96]. Several logic-based approaches have been used to define context for information integration [FDF95, GBM99, WS01]. McCarthy introduced the *ist(C,p)* predicate to disambiguate when a proposition *p* is true in context *C* [M93]. Sowa provides an overview on facts and context in [S03].

Context has been used in multiple settings: Semantic knowledge bases utilize a partial understanding of context; WordNet is such an example, where context is expressed in natural language [F98]. It has also been used to provide better algorithms for ranked answers by incorporating context into the query answering mechanism [ART06] and to improve query retrieval accuracy [STZ05]. Graphs that represent context have also been used to provide focused crawling to identify relevant topical pages on the web [DCL00]. Examples of context use in software agents, sensors and embedded devices can be found in [LS00]. Methods to model and represent context for sensor fusion using relational database model are described in [WSA02].

The problem of information integration has also attracted a lot of research work. Several approaches on schema and data integration have been presented over the years [BLN86, OS99, RB01]. There has been a significant amount of work on data integration, especially on resolving discrepancies of different data schemas using a global (mediated) schema [FLM99, LRO96, MBD05, MHH01, PGU96, RB01, RPH02]. Also, there exists work on decentralized data sharing [BLL04, DDH03, HIS03, RGJ05, TH04] and on integrating data in web-based databases [BLL04, CHZ05, DB03]. Clustering, classification and ontologies have also been used as a tool to solve semantic heterogeneity problems [DH05, JZ04, KS03, RP04, SAA04, SAK03, SBA02, ZR02, ZR04]. The discovery of semantic similarity in [FKN91] based on generalization/specialization, and positive/negative association between classes is also quite relevant, as well as discovering and ranking semantic relationships for the semantic web [AHA05, SAA04].

Our approach is relevant to the specification of relationships among database objects stored in heterogeneous database systems [GKG97, KRS99, RSK91, SK93]. We have created a methodology to derive semantic relationships among data sources based on source descriptions of participating information repositories using a metadata approach. Quite related is the work on ConceptNET, a large scale concept base that captures common sense concepts [LS04] which utilize context information.

Data Integration using Context and Semantic Networks

Context plays a significant role in answering queries which may be affected by changes in contextual information. For example, sudden changes in the environment may affect the result of a query of a scientist who is looking for the reason of the decline of fish population in a stream. If the pollution levels of the stream drastically increase due to a pollutant spillage in a nearby industrial factory, this information provides the key answer to the user's query, but only if context information is provided to the system. Note that the user may not be aware of such contextual information that has such an impact on the original query. In our example, the user may be unaware of the pollutant spillage, which is the key information and provides the exact reason for the drop in fish population. Therefore, it is quite important to use context and discover data sources that are semantically relevant and incorporate them into an enhanced query.

When context information is captured and submitted to the semantic network along with the original user query, the semantic network produces a list of additional data sources which are potentially relevant to the user's search. In our example, the semantic network would produce a list of data sources which contain pollution data.

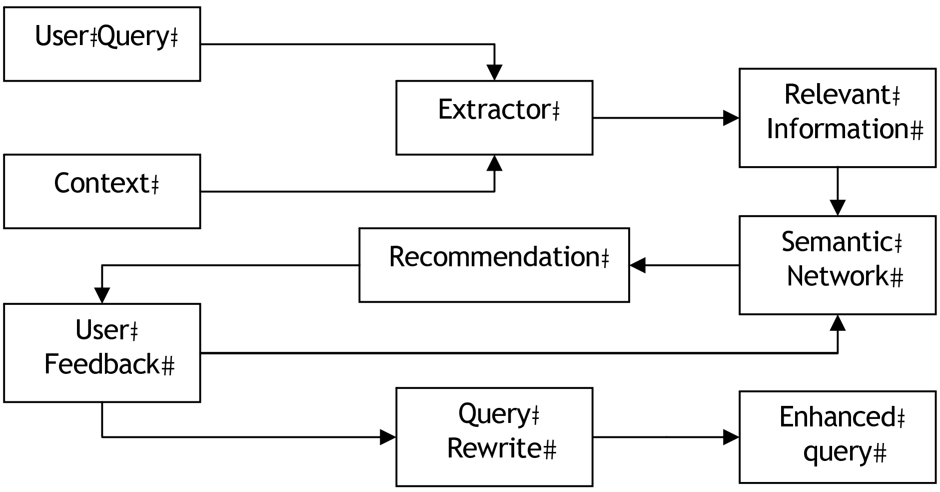

Figure 2. Enhancing a user query with context information and additional relevant semantics using a semantic network

Figure 2 illustrates the flow of a user query and the transformations that it undergoes to encapsulate context information and recommendation of additional sources, as it passes through our system. Initially both the user query and context information are submitted to the system. An extraction process retrieves relevant information from the user query and the context (for example, keywords related to the pollution data source), and presents it to the semantic network, which in turn produces a set of recommended data sources potentially interesting to the user, and related to the query. These recommendations are evaluated by the user, who has the ablility to accept or reject each one of them. At this point the user may fine-tune the search query, and/or resubmit to the semantic network and possibly provide a different threshold for relevant recommendations. This cycle may continue until the user is satisfied with the system's recommendations. Finaly, the original query will be re-written to capture the recommended data sources (pollution data source) and relevant context information. This is the final, enhanced query that will be decomposed and submitted to the participating sources.

Several software components are needed to realize the above concept and to take advantage of context information for better and more accurate semantic integration of heterogeneous resources. A simplified design of our system is presented in Figure 3, which depicts three high level components: (i) User Interface, for the interaction between users and the system, (ii) Intelligent Integration, for the enhancement and transformation of the initial user query to another one, semantically more applicable through the use of context and semantic networks, and (iii) participating Data Sources, which are various repositories of information possibly heterogeneous; we assume that each data source has a wrapper which helps facilitate interoperability between the user and the data sources.

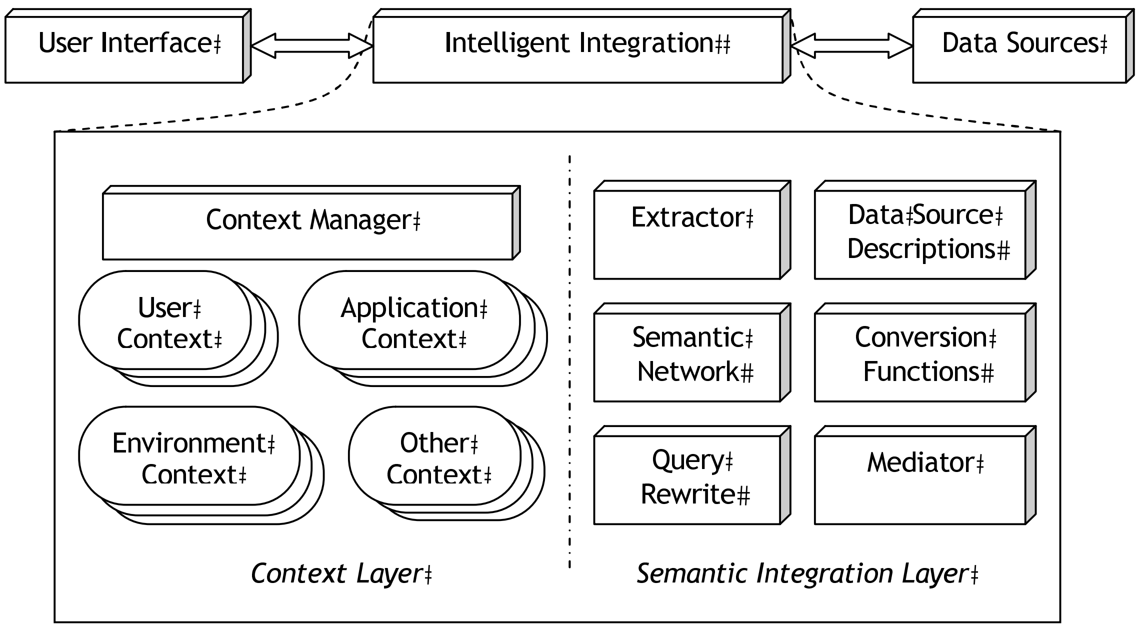

Figure 3. Major modules of the Intelligent Integration component of our system

The Intelligent Integration component is depicted in more detail. It is divided into two complementary layers: The Context Layer and the Semantic Integration Layer.

The **Context Layer** stores information on various types of context, to assist in identifying the contextual information that is relevant to the user query and thus, to supplement the query with its relevant context. It consists of the Context Manager and various types of contexts.

The *Context Manager* is the software module in charge of all context information which is collected in our system. There are several types of context in our system:

- *User Context* contains information specific to users such as personal information, recent interests, occupation, etc.
- *Application Context* contains information about the application specifics, such as type of Operating System, domain of application, information about the user's computer, etc.
- *Environment Context* collects information about environmental conditions, such as temperature, significant events and their values, etc.
- *Other Context* is used as a place holder for additional information that is collected and does not fit in any of the previous context categories.

Note that there are additional categories of context which do exist, but are not captured in our system. We acknowledge that it is unlikely to capture all possible types of context and their values in a computer system, since there will always be additional information contributing to context. We limit ourselves in collecting information about the above categories of context that are defined in our system, and we do not claim that we can capture all possible context types.

The **Semantic Integration Layer** takes advantage of the query context and utilizes the semantic network to perform the query enrichment process as described in Figure 2. It also interoperates with the participating data sources for information exchange and data integration.

- *Extractor, Semantic Network, Query Rewrite*: These are the software components that extract keywords from the user query and the relevant contexts, pass this information to the semantic network to identify additional relevant sources and perform a query rewrite to encompass all the above information into a new query (after receiving user feedback).
- *Data Source Descriptions, Conversion Functions, Mediator*: These modules assist in the submission and execution of the user's query at the data sources. As such, this layer contains descriptions about the data sources, conversion functions for mappings between different data formats, and a mediator for managing and overseeing data exchange and processing between our system and the participating data sources [W92].

The Extractor is a simple component that extracts keywords from the user query and the relevant contexts. The semantic network recommends additional relevant data sources (for details see [CGK07]). The Query Rewrite component uses a simple algorithm to transform the initial query into an enhanced one as follows: First, the relevant keywords are extracted from the contexts and the semantic network. Then each data source (corresponding to a relevant keyword) is added in the FROM clause of the original user query, and the necessary joins are included in the WHERE clause. The SELECT clause is also augmented with the joining attributes and all attributes that are common in the used contexts. The enhanced query is provided to the Mediator for further processing. The Mediator decomposes the query into subqueries, submits them to the participating data sources (information repositories), taking care of possible incompatibilities using the conversion functions, collects the subquery results from the data sources, and performs a post-processing step to put together the various subquery results and present the final result to the user.

Implementation and Evaluation

We have implemented a prototype system to demonstrate the validity of our methodology. We used environmental domain data to exhibit the improvement we gained by running the transformed user query through the described techniques, and compared it with the initial user query without passing it through the Intelligent Integration component.

Our experimental setting contains descriptions of several data sources. The first data souce holds information related to fish species, their population, and the location they live in (stream, lake, ocean, etc.). The second data source contains information about the environment, such as temperature, pollution, pollutants and how to treat contaminated areas. We also collect information about the users, such as name, profession, hobbies, and most importantly a history of their past queries. Several components of the above information are also part of the context: Data related to user profiles (hobbies, past queries), and environment (pollutant, pollution, and temperatures) are identified as User Context and Environment Context, respectively. For this particular experiment we did not use Application or Other Context.

Picture the following scenario: A marine biologist is interested in getting information about the population of the Atlantic salmon in a specific geographic area. The user accesses the Fish Information data source and submits a query. The result of this query shows a big drop in salmon population over time: the fish count decreased from about 500 down to 182 - a significant fall within three weeks. Puzzled by these numbers, the marine biologist is intrigued to find the reasons for this decline. Note that this query was submitted to a single data source without enhancement, and the only information that is revealed is the decline in salmon population.

Next, the user submits the same query to our prototype. The Intelligent Integration component consults the User and Environment Contexts, and the Extractor identifies additional keywords (“temperature”, “location”) from these contexts. These keywords along with the original keyword “fish” are submitted to the semantic network, which in turn recommends the pollution data source to be incorporated into the user query since it is semantically related and indirectly connected to the Fish Information source. The marine biologist concurs and the initial query is being transformed into an enhanced one by the Query Rewrite component. Now the query spans both data sources, fish and pollution.

This time, the transformed query retrieves not only information about salmon population, but pollution levels over the same time period.

(a)	(b)
`select name, count_date, count from fish, fish_count where fish.name = 'Salmon' and fish.id = fish_count.id order by count_date;`	`select f.name, fc.count_date, fc.count, p.location_name, p.pollution_level from fish f, fish_count fc, pollution p, water_container w where f.name = 'Salmon' and f.id = fc.id and f.loc_id = w.id and w.name = p.location_name order by count_date;`

Figure 4. (a) Original user query, and (b) Enhanced user query

Figure 4 shows simplified versions of the original user query and the enhanced query as was transformed by the Query Rewrite component after consulting the contexts and the recommendations of the semantic network.

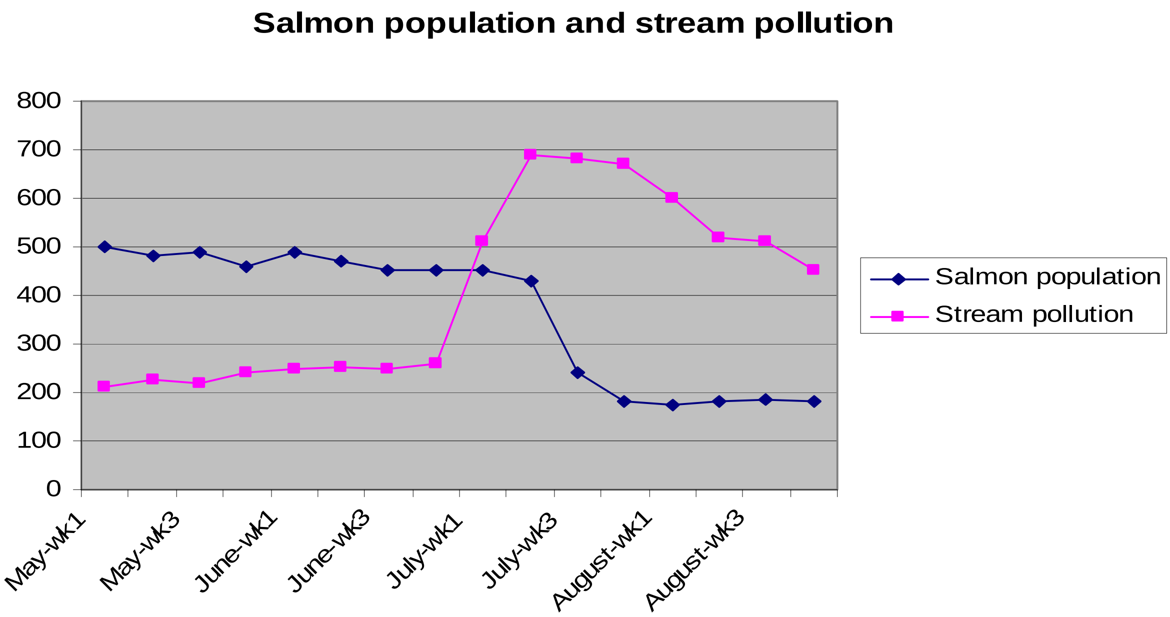

Figure 5. Salmon population and water pollution measures over a time period

Figure 5 illustrates partial results of the enhanced user query as a plot. It identifies that the decrease in salmon population follows an increase in pollution levels over the same time period. The two plotted lines together provide a satisfactory explanation to the marine biologist, who now knows that the increased pollution levels caused the decline in fish counts. We performed a number of additional scenarios and all had similar results.

Conclusions and Further Work

In this paper we presented a methodology, which exploits context information and semantic networks to assist in the semantic data integration of heterogeneous information systems. The main contribution of this paper is the combined use of both context and semantic network techniques to recommend to users additional data sources that are relevant to the original query and convey context information. The original user query is transformed into an enhanced query, which succintly conveys the intention of the user; thus, when submitted to the data sources it returns results that better match the initial user intent, as it incorporates information about context and semantics. We implemented a prototype system and performed a experiments which verified our claims.

We plan to expand our work and apply statistical models to context and analyze semantic networks through statistics. In addition, we plan to monitor changes in the context and dynamically update semantic networks using Even-Condition-Action (ECA) rules using the Oracle DBMS which provides an Expression Filter facility to automatically implement ECA rules [O05, TGP05].

References

[AHA05] Aleman-Meza, B.; Halaschek-Wiener, C.; Arpinar, I. B.; Ramakrishnan, C.; Sheth, A. P.: "Ranking Complex Relationships on the Semantic Web," *IEEE Internet Computing*, pp. 37-44, 2005.

[ART06] Agrawal, R.; Rantzau, R.; Terzi, E.: "Context-Sensitive Ranking," ACM SIGMOD International Conference on Management of data, Chicago, IL, USA, 2006.

[B96] Brezillon, P.: "Context in Human Machine Problem Solving: A Survey," LAFORIA, Research Report 1996.

[BB05] Bazire, M.; Brezillon, P.: "Understanding Context before Using It," CONTEXT 2005. In Lecture Notes in Artificial Intelligence, 2005.

[BLL04] Bowers, S.; Lin, K.; Ludascher, B.: "On Integrating Scientific Resources through Semantic Registration," Scientific and Statistical Database Management, 2004.

[BLN86] Batini, C.; Lenzerini, M.; Navathe, S.: "A Comparative Analysis of Methodologies for Database Schema Integration," ACM Computing Surveys, vol. 18, pp. 323--364, 1986.

[CGK07] Chen, Z.; Gangopadhyay, A.; Karabatis, G.; McGuire, M.; Welty, C.: "Semantic Integration and Knowledge Discovery for Environmental Research," Journal of Database Management, vol. 18, pp. 43-67, 2007.

[CHZ05] Chang, K. C.-C.; He, B.; Zhang, Z.: "Toward Large Scale Integration: Building a Metaquerier over Databases on the Web," CIDR, 2005.

[DB03] Dispensa, J. M.; Brulle, R. J.: "The Sprawling Frontier: Politics of Watershed Management," Submitted to Rural Sociology, 2003.

[DCL00] Diligenti, M.; Coetzee, F.; Lawrence, S.; Giles, C. L.; Gori, M.: "Focused Crawling Using Context Graphs," 26th International Conference on Very Large Data Bases, Cairo, Egypt, 2000.

[DDH03] Doan, A.; Domingos, P.; Halevy, A. Y.: "Learning to Match the Schemas of Data Sources: A Multistrategy Approach.," Machine Learning, vol. 50, pp. 279-301, 2003.

[DH05] Doan, A.; Halevy, A. Y.: "Semantic-Integration Research in the Database Community," AI Mag., vol. 26, pp. 83-94, 2005.

[F98] Fellbaum, C.: Wordnet: An Electronic Lexical Database. Cambridge, MA, USA: MIT Press, 1998.

[FDF95] Farquhar, A.; Dappert, A.; Fikes, R.; Pratt, W.: "Integrating Information Sources Using Context Logic," AAAI Spring Symposium on Information Gathering from Distributed Heterogeneous Environments, 1995.

[FKN91] Fankhauser, P.; Kracker, M.; Neuhold, E. J.: "Semantic Vs. Structural Resemblance of Classes," SIGMOD Record, vol. 20, pp. 59-63, 1991.

[FLM99] Friedman, M.; Levy, A.; Millstein, T.: "Navigational Plans for Data Integration," AAAI/IAAI, 1999.

[GBM99] Goh, C. H.; Bressan, S.; Madnick, S.; Siegel, M.: "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information," ACM Transactions on Information Systems, vol. 17, pp. 270-293, 1999.

[GKG97] Georgakopoulos, D.; Karabatis, G.; Gantimahapatruni, S.: "Specification and Management of Interdependent Data in Operational Systems and Data Warehouses," Distributed and Parallel Databases, vol. 5, pp. 121-166, 1997.

[HIS03] Halevy, A. Y.; Ives, Z. G.; Suciu, D.; Tatarinov, I.: "Schema Mediation in Peer Data Management Systems," IEEE International Conference on Data Engineering (ICDE), 2003.

[JZ04] Jain, H.; Zhao, H.: "Federating Heterogeneous Information Systems Using Web Services and Ontologies," Tenth Americas Conference on Information Systems, New York, NY, 2004.

[KRS99] Karabatis, G.; Rusinkiewicz, M.; Sheth, A.: "Interdependent Database Systems," in Management of Heterogeneous and Autonomous Database Systems. San Francisco, CA: Morgan-Kaufmann, 1999, pp. 217-252.

[KS03] Kalfoglou, Y.; Schorlemmer, M.: "Ontology Mapping: The State of the Art," Knowledge Engineering Review, vol. 18, pp. 1-31, 2003.

[KS96] Kashyap, V.; Sheth, A.: "Semantic and Schematic Similarities between Database Objects: A Context-Based Approach," The VLDB Journal, vol. 5, pp. 276-304, 1996.

[LRO96] Levy, A. Y.; Rajaraman, A.; Ordille, J. J.: "Querying Heterogeneous Information Sources Using Source Descriptions," VLDB Conference, 1996.

[LS00] Lieberman, H.; Selker, T.: "Out of Context: Computer Systems That Adapt to, and Learn from, Context," IBM Systems Journal, vol. 39, pp. 617-632, 2000.

[LS04] Liu, H.; Singh, P.: "Conceptnet: A Practical Commonsense Reasoning Toolkit," BT Technology Journal, vol. 22, pp. 211-226, 2004.

[M61] Masterman, M.: "Semantic Message Detection for Machine Translation, Using an Interlingua," NPL, pp. 438-475, 1961.

[M93] McCarthy, J.: "Notes on Formalizing Context," International Joint Conference on Artificial Intelligence (IJCAI), Chambéry, France, 1993.

[MBD05] Madhavan, J.; Bernstein, P. A.; Doan, A.; Halevy, A. Y.: "Corpus-Based Schema Matching," IEEE International Conference on Data Engineering (ICDE), 2005.

[MHH01] Miller, R. J.; Hernandez, M. A.; Haas, L. M.; Yan, L.; Ho, C. T. H.; Fagin, R.; Popa, L.: "The Clio Project: Managing Heterogeneity.," SIGMOD Record, vol. 30, 2001.

[O05] Oracle: "Application Developer's Guide – Rules Manager & Expression Filter 10g Release 2," Oracle Corporation Dec 2005.

[OS99] Ouksel, A.; Sheth, A. P.: "Special Issue on Semantic Interoperability in Global Information Systems," SIGMOD Record, vol. 28, 1999.

[P00] Pearl, J.: Causality: Models, Reasoning, and Inference: Cambridge University Press, 2000.

[P88] Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference: Morgan Kaufmann, 1988.

[PB99] Pomerol, J.-C.; Brezillon, P.: "Dynamics between Contextual Knowledge and Proceduralized Context," Modeling and Using Context (CONTEXT-99). In Lecture Notes in Artificial Intelligence, 1999.

[PGU96] Papakonstantinou, Y.; Garcia-Molina, H.; Ullman, J.: "Medmaker:A Mediation System Based on Declarative Specifications," IEEE International Conference on Data Engineering (ICDE), 1996.

[R94] Rice, J. A.: Mathematical Statistics and Data Analysis: Duxbury Press, 1994.

[RB01] Rahm, E.; Bernstein, P. A.: "A Survey of Approaches to Automatic Schema Matching," VLDB Journal, vol. 10, 2001.

[RGJ05] Rodriguez-Gianolli, P.; Garzetti, M.; Jiang, L.; Kementsietsidis, A.; Kiringa, I.; Masud, M.; Miller, R.; Mylopoulos, J.: "Data Sharing in the Hyperion Peer Database System," International Conference on Very Large Databases (VLDB), 2005.

[RP04] Ram, S.; Park, J.: "Semantic Conflict Resolution Ontology (Scrol): An Ontology for Detecting and Resolving Data and Schema-Level Semantic Conflicts," IEEE Transactions on Knowledge and Data Engineering, vol. 16, pp. 189-202, 2004.

[RPH02] Ram, S.; Park, J.; Hwang, Y.: "Cream: A Mediator Based Environment for Modeling and Accessing Distributed Information on the Web," British National Conference on Databases (BNCOD), 2002.

[RSK91] Rusinkiewicz, M.; Sheth, A.; Karabatis, G.: "Specifying Interdatabase Dependencies in a Multidatabase Environment," IEEE Computer, vol. 24, pp. 46-53, 1991.

[S03] Sowa, J. F.: "Laws, Facts, and Contexts: Foundations for Multimodal Reasoning," in Knowledge Contributors, Hendricks, V. F., Jorgensen, K. F., and Pedersen, S. A., Eds. Dordrecht: Kluwer Academic Publishers, 2003, pp. 145-184.

[S06] Sowa, J. F.: "Semantic Networks http://www.jfsowa.com/pubs/semnet.htm."

[S92] Sowa, J. F.: "Semantic Networks," in Encyclopedia of Artificial Intelligence, Shapiro, S. C., Ed., Second ed. New York: Wiley, 1992, pp. 1493-1511.

[SAA04] Sheth, A.; Aleman-Meza, B.; Arpinar, I. B.; Bertram, C.; Warke, Y.; Ramakrishanan, C.; Halaschek, C.; Anyanwu, K.; Avant, D.; Arpinar, F. S.; Kochut, K.: "Semantic Association Identification and Knowledge Discovery for National Security Applications," Journal of Database Management, vol. 16, 2004.

[SAK03] Sheth, A.; Arpinar, I. B.; Kashyap, V.: "Relationships at the Heart of Semantic Web: Modeling, Discovering, and Exploiting Complex Semantic Relationships," in Enhancing the Power of the Internet Studies in Fuzziness and Soft Computing, Nikravesh, M., Azvin, B., Yager, R., and Zadeh, L. A., Eds.: Springer-Verlag, 2003.

[SBA02] Sheth, A.; Bertram, C.; Avant, D.; Hammond, B.; Kochut, K.; Warke, Y.: "Managing Semantic Content for the Web," IEEE Internet Computing, vol. 6, pp. 80-87, 2002.

[SK93] Sheth, A.; Karabatis, G.: "Multidatabase Interdependencies in Industry," ACM SIGMOD International Conference on Management of Data, Washington DC, 1993.

[STZ05] Shen, X.; Tan, B.; Zhai, C.: "Context-Sensitive Information Retrieval Using Implicit Feedback," 28th international ACM SIGIR conference on Research and development in information retrieval, Salvador, Brazil, 2005.

[TGP05] Thome, B.; Gawlick, D.; Pratt, M.: "Event Processing with an Oracle Database," ACM SIGMOD International Conference on Management of Data, Baltimore, Maryland, 2005.

[TH04] Tatarinov, I.; Halevy, A. Y.: "Efficient Query Reformulation in Peer-Data Management Systems.," ACM SIGMOD International Conference on Management of Data, 2004.

[W92] Wiederhold, G.: "Mediators in the Architecture of Future Information Systems," IEEE Computer, vol. 25, pp. 38-49, 1992.

[WS01] Wache, H.; Stuckenschmidt, H.: "Practical Context Transformation for Information System Interoperability," International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2001). Lecture Notes in Artificial Intelligence, 2001.

[WSA02] Wu, H.; Siegel, M.; Ablay, S.: "Sensor Fusion for Context Understanding," 19th IEEE Instrument and Measurement Technology Conference, Anchorage, AK, USA, 2002.

[ZR02] Zhao, H.; Ram, S.: "Applying Classification Techniques in Semantic Integration of Heterogeneous Data Sources," Eighth Americas Conference on Information Systems, Dallas, TX, 2002.

[ZR04] Zhao, H.; Ram, S.: "Clustering Schema Elements for Semantic Integration of Heterogeneous Data Sources," Journal of Database Management, vol. 15, pp. 88-106, 2004.

[ZRG05] Zadorozhny, V.; Raschid, L.; Gal, A.; Ye, Q.; Murthy, H.: "Using Non-Random Associations for Predicting Latency in Wans," International Conference on Web Information Systems Engineering (WISE), New York, NY, USA, 2005.

Biography

George Karabatis is an Assistant Professor of Information Systems at the University of Maryland, Baltimore County (UMBC). He holds degrees in Computer Science (Ph.D. and M.S.) and Mathematics (B.S.). Before joining UMBC he was a Research Scientist at Telcordia Technologies (formerly Bellcore) working on database research for the telecommunications industry. His current research interests include semantic information integration and applications for mobile handheld devices. He is a member of AIS, ACM, IEEE, IEEE-CS.