Challenges in Making Context Interchange Part of the Semantic Web

Michael Siegel (msiegel@mit.edu)
Stuart Madnick(smadnick@mit.edu)
Aykut Firat (aykut@mit.edu)
Allen Moulton(amoulton@mit.edu)
Hongwei Zhu (mrzhu@mit.edu)

MIT Sloan School of Management
30 Wadsworth Street, Cambridge, MA 02142,
USA
http://context2.mit.edu/coin

1. Introduction

The popularity and growth of the Web have dramatically increased the number of information sources available for use and the opportunity for important new information-intensive applications (e.g., massive data warehouses, integrated supply chain management, global risk management, in-transit visibility). Unfortunately, there are significant challenges to be overcome regarding *data interpretation*. Specifically, the existence of heterogeneous contexts, whereby each source of information and potential receiver of that information may operate with a different context, leading to large-scale semantic heterogeneity.

A context is the collection of implicit assumptions about the meaning of data. As a simple example, whereas most US universities grade on a 4.0 scale, MIT uses a 5.0 scale – posing a problem if one is comparing student GPA's. Another typical example might be the extraction of price information from the Web: but is the price in Dollars or Yen (If dollars, is it US dollars or Hong Kong dollars), does it include taxes, does it include shipping, etc. – and does that match the receiver's assumptions?

Contextual issues can be much more complex in other situations. For example, the meaning of "net sales" may vary – with "excise taxes" included for government reporting purposes in one context, but excluded for security analysis purposes in another. Also, one context may use information for a fiscal year as reported by the company, while another may use a standardized fiscal year to make all companies comparable. Furthermore, there may be multiple users that might want an answer to such a question, each with their own desired meaning (user context).

This context knowledge is often widely distributed within and across organizations. Solutions adopted to achieve interoperability must be scaleable and extensible. Thus, it is important to support the acquisition, organization, and effective intelligent usage of distributed context knowledge.

The COntext INterchange (COIN) System has been designed and implemented as a prototype at MIT. The prototype provides for a systematic representation and automatic processing of data semantics. Instead of explicitly capturing semantic conflicts, the COIN approach records data semantics declaratively and uses a mediation engine to detect all conflicts, which are reconciled by rewriting user queries to incorporate conversions that can be defined either internally or remotely on the network. This approach provides great extensibility. We refer readers to [1,2] for a formal description of the COIN approach.

2. Recent Developments in COIN

Recent developments by Firat [10] have provided a clear definition of concepts such as *context, conversion function, and ontology*. His work also resolved issues in *equational ontological conflicts* (EOC) that refer to the heterogeneity in the way data items are calculated from other data items in terms of definitional equations. Firat along with others at MIT have developed an approach to merging independently developed, ontology based COIN *applications*. Finally, there have been significant developments in providing for semantic integration using COIN on the Semantic Web. Specifically we have developed ways to make the context mediation approach compatible with web protocols (as in web services) and web-oriented representation languages such as RDF and OWL)[4,5].

We have demonstrated these new capabilities in a number of application domains, such as financial services [6], online shopping [9], disaster relief efforts [3], and corporate householding knowledge engineering [8]. We have also constructed larger applications by combining ontologies and context definitions from existing applications, such as an airfare aggregator and a car rental shopper combined into a travel planner (see demos at our website http://context2.mit.edu/coin/demos).

Efforts are also underway to use COIN framework as a cost effective method for resolving semantic ambiguities and differences to support semantic interoperability across multiple overlapping standards in the financial industry [7].

3. Making Context Mediation Ubiquitous on the Web: The Challenges

Our approach to semantic integration is data-oriented. As such, our goals are far more focused than many other visions of the potential for the Semantic Web. As a result, we are able to treat context interchange problems inherent in the Semantic Web in a tractable manner. For example, we

have a specific approach to merging ontologies that supports the merging of applications. This merging raises many of the issues that others have looked at but is nicely tied to the data requirements for new applications and focused on providing the context information needed to resolve semantic differences.

This focus on context knowledge and data integration has allowed us to make significant progress, however, challenges exist in making such an approach scalable, maintainable and usable in an open environment. We conclude this position paper with a number of these issues:

1. **Gathering, Representing and Maintaining Context Knowledge for Unknown Tasks** – Context Interchange capabilities have been used for specific applications. Though the semantic integration can be done at run-time for such an application, ad-hoc environments without predefined schemas and context knowledge will be more difficult to manage.
2. **Designing Ontologies to Include Context Knowledge** – We have developed ontology to support context knowledge. We have extended ontologies developed in RDF to include modifiers and other context information. However, we expect a wide range of ontology languages and representations. Context information must either be easily extracted from these ontologies or added through the use of context-authoring tools as developed on this project.
3. **Making Context Mediation Executable in non-SQL like environments** – We have taken a distinctly database-like approach to semantic integration on the Web. We developed data extraction tools that gather semi-structured data based on SQL queries issues to Web pages (along with structured data). Methods must be developed to include mediation in for other data representations.
4. **Automatically Gathering, Generating and Maintaining Context Knowledge** – Tools are needed to automatically assemble and maintain context knowledge.
5. **Complex Context Issues** – There remain a number of complex context issues related to temporal context, equational context, and partially resolvable context conflicts.

References

1. Goh, C.H. (1997) Representing and Reasoning about Semantic Conflicts in Heterogeneous Information Systems, PhD Thesis, MIT Sloan School of Management
2. Goh, C.H., Bressan, S., Madnick, S., Siegel, S. (1999) "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information", *ACM Transactions on Information Systems,* **17**(3), 270-293
3. Lee, P.W. (2003) Metadata Representation and Management for Context Mediation, MIT EECS Master's Thesis.
4. Liburd, S. (2003) Conceptual Mapping and Structural Conversions between COIN and RDF, technical report, MIT Sloan School of Management.
5. Manola, F. Miller, E. (2003) "RDF Primer", W3C working draft.
6. Firat, A., Madnick, S., Grosof, B. (2002) Financial Information Integration in the Presence of Equational Ontological Conflicts, *12th WITS*, December 14-15, Barcelona, Spain.
7. Moulton A., Siegel, M.D., Madnick, S.E. (2003) Potential for Cost Savings from Application of Context Mediation Technology to Problems of Standards Interoperability in the Financial Services Industry, internal report, MIT Sloan School of Management.
8. Xian, X. (2003) Corporate Householding Knowledge Engineering and Processing using Extended COIN, MIT EECS Master's Thesis.
9. Zhu, H. Madnick, S., Siegel, M. (2002) "Global Comparison Aggregation Services", *1st Workshop on E-Business*, December 14-15, Barcelona, Spain.
10. Firat, A. (2003) "Information Integration Using Contextual Knowledge and Ontology Merging" PhD Thesis, MIT Sloan School of Management.