Chapter 6.   Conclusions and Future Research Directions

6.1. Summary

Heterogeneous information integration on the Web involves a number of new techniques. This includes mechanisms for information encoding and manipulation (e.g. XML, RDF, XSLT), and ontology construction and reasoning (e.g. RDFS, DAML+OIL, OWL). In order to manage and use information more effectively within the enterprise, a benchmark used to evaluate the mechanism of heterogeneous information integration is needed. In this research, we have developed the XML and ontology benchmark workload model in heterogeneous information integration, and built a workload generation prototype. We have reviewed the XML and ontology related literature to motivate the design of the workload model. The objective of this research is to develop a workload model to test whether the heterogeneous information integration system under EB environment can overcome the diverse formats of content and derive meaning from this content. In order to apply the workload model to different scenarios easier, it is designed in generic constructs. Finally, we validate the research model through the prototype implementation. The results in this research include:

- Collecting and reviewing the literature on XML standardization, XML benchmarks, ontology standardization, and ontology related benchmarks. Identifying major requirements for a XML or ontology benchmark.

- Developing a generic XML and ontology benchmark workload model in heterogeneous information integration. The workload model consists of XML and ontology data model and query model according to the generic constructs and constrains requirements. Also, a control model is created to set up the benchmark environment.

- Implementing a workload generation prototype based on the workload model in this research. The prototype illustrates the feasibility and validity of the research model.

6.2. Future Research Directions

This research only built a simple prototype of the XML and ontology benchmark workload model of heterogeneous information integration. It still needs more effort to expand its capabilities. We expect this work to continue and evolve in the future. Future research directions include:

- Enhancing the ontology query model. The development of an ontological standard presents many opportunities and challenges. New reasoning tasks may arise in the future. Retrieval (instances of a concept) and realization (most specific class of instance) may not be sufficient. In order to make the ontology query model more comprehensive, further study to keep track of ontology progression is needed.

- Improving the complexity factors of the XML query model. The complexity factors we analyze in the XML query model are still too rough. Each query type can be analyzed more carefully to refine the query model.

- Implementing various data distributions. In this research, only uniform distribution is implemented. It cannot evaluate performance under different distributions. Implementation of diverse data distributions will become a user requirement.

- Applying the workload model to other applications. Ontology and XML are complementary technologies, and there are other applications that can apply. In this research, we assume the heterogeneous information integration system is used on Intranets, such as enterprise information integration (EII), electronic business (EB), and enterprise application integration (EAI). There are other applications between enterprises that may need to integrate heterogeneous information, such as

business-to-business integration (B2Bi), collaborative commerce (C-Commerce), and electronic commerce (EC). We can modify the workload model of this research to create other benchmarks that are based on XML and ontology with different characteristics.