# Chapter 5 Research Discussions and Limitations

In this research, the results are analyzed with managerial and technical implications. We intend to provide a discussion and a reflection on the work and model we have developed. The research limitations are also described.

## 5.1. Research Implications

Due to the speedy development of Information Technology and the ubiquitous nature of the Internet, the issue of heterogeneous information integration becomes more and more important. Therefore, to achieve interoperability, scalability and portability between multiple heterogeneous information sources and provide users a more efficient way to access the different kinds of information sources become the initial motivation of this research.

There have been many works focusing on the issue of heterogeneous information integration before. They use object model to serve the common data model, which is to be the unified view of the underlying heterogeneous information sources. However, using object model to tackle such problems faces the insufficient capability for semantic heterogeneity problems.

In the light of such problems, this research tries to find a solution and propose a method of HII with schema and ontology assisted. We choose XML to be the common object model because it has been a standard format to exchange information across the Internet. However, XML can also be viewed as a kind of object model. While XML can indeed establish structural interoperability between different information

sources on the Web, it can barely reach the semantic interoperability over EB settings. To improve this shortcoming, ontology is requested here. Therefore, we improve the shortcomings of adopting XML-based HII method in the aspect of semantic interoperability. There still some research adopted pure ontology-based HII method. But we discover that adopting such method would somehow encounter the shortcoming of information losses due to the inference processing. And the linking between the ontology and local sources is more complex than object model. Besides, ontology does not yet have a uniform standard to follow. It is unfavorable for the scalability of HII. Furthermore, perfoming HII in EB environment may require higher standard over the accurancy of the requested information which is an unfavorable facor of adopting pure ontology-based HII method. Therefore, we develop an XML based with ontology assisted HII method in order to gain better interoperability over the heterogeneous information sources and avoid shortcomings of the traditional HII methods.

We implement a prototype to perform heterogeneous information integration with schema and ontology assisted for realizing the research method we propose. Users can access a number of heterogeneous information sources through a systematic way by this prototype to gain the information without inconsistency and contradictory. Furthermore, this prototype hides the underlying complexity from users. So the process of accessing information sources becomes more effective and efficient. Hence, summarizing the research method in Chapter 3 and system implementation in Chapter 4, we give some discussions of our research as follows:

1. Providing generic construct orientation not ad hoc mapping to generate global schema.

Creating the unified view, global schema, of the underlying information sources is important for the task of heterogeneous information integration especially when using the mediator/wrapper architecture. Users interact with multiple heterogeneous information sources through the global schema. But after reviewing the previous works on this issue, we discover that the creation of the global schema in the previous research almost needs ad hoc exertion. It requires human experts to take responsibility for specifying the needed global schema. However, we think that does not fit in with cost-effectiveness and might be application- and expert-dependent. As such, for this research we propose a generic construct orientation not an ad hoc mapping method to generate the global schema in order to achieve the application independence that is important to web-based HII in EB to reach the higher interoperability, scalability, and portability.

2. Efficient query reformulation by adopting GAV approach to specify the mapping between global schema and local source schema.

For this research, we choose GAV approach to establish the mapping between global schema and local source schema because easy and efficient query reformulation is our prior consideration. However, the shortcoming of the GAV approach is that it is hard to perform the evolution process. When the local source schema changes or new information source is added or the existing information source is removed, the system faces difficulty in updating the mapping between global schema and local schema. As such, it limits the scalability of the system for the future. In contrast, the LAV approach has the advantage of the evolution process, but it is more complex against the query reformulation process. So it may limit the capability of the system to reach the complete interoperability of both structure and semantics between heterogeneous information sources, which is a prior consideration

of this research. Consequently, we still choose GAV approach to establish the mapping between global schema and local source schema, though the system may meet the difficulty of evolution in the future.

3. The shortcoming of GAV approach is evolution complexity, so GAV approach is best used in stable environment.

As described above, the shortcoming of the GAV approach is evolution complexity. So we suggest that our method of heterogeneous information integration is best used in a stable environment. Since a prerequisite of this research is to perform the web-based HII in EB in which the environment would not change frequently, we think it is still applicable to solve the problem.

4. Using XML and XQuery standards as interface; and mainstream data sources including relational model (RM), object model (OM), and web pages.

In this research, we use XQuery standards as interface because we use XML as the common data model in the method. In the system implementation, we adopt three mainstream information sources including relational model (RM), object model (OM) and web pages as representative heterogeneous information sources. Through this implementation, we think it can prove that our method can be applied to main kinds of heterogeneous information sources. In other words, our method could bear high heterogeneity and reach the interoperability between them, which is our goal.

5. Global schema with ontology really helps query with reasoning which in turn structural and semantic interoperability but working together now is primitive and domain-specific.

Since we use XML as the format for global schema, it can hardly catch complete relationships. For example, it is hard to represent the relationship of subsumption and overlapping. Due to a lack of real world cases, the current prototype does not cover all kinds of the relationships. The expressive power of the ontology language is another factor of this restriction, so the global schema works together with the ontology to reach the structural and semantic interoperability now is primitive and domain specific.

## 5.2. Research Limitations

Due to time and technical constraints, this research still has room to improve. In this sub-section, we describe several limitations of this research as follows:

1. Part I of the creation process of the global schema, i.e. to transform local information sources structure into generic XML Schema, is done. But Part II of the creation process of conflicts and correspondences identification is still manual and expert-dependent.

2. Because the local information source structure may be quite different from the generic XML Schema in essence, the transformation from the local information source data model to the generic XML Schema may meet some exceptions that lead to the incomplete transformation. When meeting such problems, we treat them as the special or even exceptional case by creating some annotations. Afterward, we can specify some rules to process those exceptional cases when implementing the transformation method.

3. Due to the heterogeneity of the underlying information sources, query languages

used in different sources are also quite different from the XQuery in essence. So translating the user query in XQuery expression into query languages used in local sources may also meet some exceptions. For example, “path traversal” specifying in XQuery may hard to map to SQL. Again we treat those exceptions as special cases. We need markup those exceptions and write additional rules when programming for processing the exceptions.

4. In this research, we adopt the GAV approach to specify the mapping between global schema and local source schema in order to gain the interoperability between multiple heterogeneous sources. However, due to the complexity of the evolution process of the GAV approach, a stable and static evolution is assumed.

5. Due to time restrictions, we only implement a primitive parser for XQuery in terms of syntax in the prototype system.