Chapter 6

Integration of Multiple Biological Databases

6.1 Introduction

In chapter 5 we have described our proposed system, INQUES, for querying different biological databases expressed in XML format. Using the same architecture for integration of biological databases it is possible to integrate multiple biological databases. As discussed in chapter 1 there are various kinds of integration approaches used for integrating biological databases like warehouse integration, mediator-based integration, navigational integration and federated database approach. The advantages and drawbacks of these approaches have also been discussed in chapter 2.

This chapter presents integration of multiple biological databases using INQUES whereby we have been able to overcome the major drawbacks of the existing systems. Through this system it is possible to integrate various biological databases online.

6.2 Integration of Multiple Biological Databases Expressed in XML Format

In chapter 3 and 4 it has been shown that XML DTDs can be transformed/mapped to BAO terms. So BAO terms can be considered as common terms to link various biological databases expressed in XML format. The query input is given in terms of BAO to integrate various biological databases. This input value is used by the system to search different XML databases, and wherever a match is found the system stores the corresponding entries related to that particular biological database. If this match is found in
more than one biological database then the system appends the related entries of that database to the previous result. Here the entries that would be found will be originally in terms of XML format corresponding to their respective biological databases. A query translation is done at this stage to translate these entries into BAO terms. The final query result is displayed in terms of BAO.

This can be explained as follows.

Lets say that the query input is in term of ontology like, \textit{access-no} = "AE000795", and the desired output format is in terms of BAO. The system searches the corresponding term in different XML DTDs. If the entry corresponding to \textit{access-no} is found in AGAVE, for example, then the biological data is retrieved in the AGAVE format. Then the query result is translated into ontology term through the query translation using table 5.4. Now the system searches another XML DTD for the same entry. If the term corresponding to \textit{access-no} is found in that DTD, then it retrieves the result in that particular XML DTD format and translates it to ontology terms. This result is appended to the previous query result, and so on.

Here the XPath Query would be,

\[
/protein-id-info \text{[access-no = 'AE000795']}
\]

From the table the translation for AGAVE would be

\[
/db_id[ \text{id = 'AE000795'}]
\]

and the result would be in AGAVE format.

Similarly we have queried other data sources and the result is appended to the previous result.
6.3 Implementation

INQUES has been developed for integration and querying of multiple biological databases expressed in XML format. Here we achieve integration by querying the biological data in terms of BAO and storing the corresponding query result also in terms of BAO as discussed in section 6.2.

Figure 6.1 Query example: Input: BAO, Output: BAO

In figure 6.1 the query input is in terms of BAO, and the query output is also in terms of BAO. The query result is shown in figure 6.2.
Figure 6.2 Query result of example shown in figure 6.1

6.4 Conclusion

In this chapter we have used INQUES for integration of biological databases. This system overcomes the major drawbacks of the existing systems. Through this system we have been able to integrate various biological databases online.