

# XML資料整合與關聯法則分析及其在蛋白質資料庫之應用

張俊堂, 劉志俊

資訊工程學系

資訊學院

ccliu@chu.edu.tw

## 摘要

As the fast increasing of XML data, data mining techniques can discover valuable rules behind these large amounts of data. There are many biological databases which provide XML data for data interchanging purpose. If we can integrate these biological XML data, more meaningful information can be derived and their applications can be expanded. Therefore, in this paper, we propose a technique to integrate protein XML data which come from different sources. Then a method for mining integrated XML trees is provided. The proposed method can discover association rules among hierarchical structures of the integrated XML trees. The semantics and syntaxes for the XML association rules are also discussed.

關鍵字：XML, Data Mining, Association Rule, Protein Databases