Project Proposal

Study of Association Rule Mining in XML data

Balaji Ganesan (balaji*)
Richard Nikhil Martin (rnmartin*)
Satya Narayan Thiruvallur Selvakumar (satya*)

Aim:

We wish to study and implement Association rule mining algorithms in XML data and compare its performance against other such implementations.

Motivation:

Association rule mining is a form of data mining in which the goal is to discover relationships between attributes in large data sets. We find several implementations of Association Rule mining in the relational database field. However such implementations in the XML data are still limited. We ourselves could find only one such implementation using the Apriori algorithm.

Project goals:

1. Study the basics of Association rule mining and the various algorithms used for this in the relational database field.
2. Implement one of these algorithms (apart from Apriori) for XML data.
3. Compare the performance of our implementation with the existing literature.

References:

1. Daniele Braga, Alessandro Campi, Stefeno Ceri, Mika Klemettinen, and PierLuca Lanzi. Discovering Interesting Information in XML Data with Association Rules. In Proceedings of the 2003 ACM Symposium on Applied Computing, pages 450–454, Melbourne, Florida, 2003.

2. Jacky W W Wan, Gillian Dobbie**.** Mining association rules from XML data using XQuery. In Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation - Volume 32 2004.