# PolyULogo

# COMP5323 Web Database Technologies and Applications

Research Paper Review Report

# “Generating Efficient Execution Plans for Vertically Partitioned

# XML Databases”

# ***VLDB 2011 -***

# ***37th International Conference on Very Large Data Bases***

# Group Members:

QING Pei, Edward (11500811g)

LO Wing Yi, Wing (11523479g)

SHAO Shuai, Philip (11552402g)

**Paper Summary**

Title : Generating Efficient Execution Plans for Vertically Partitioned XML Databases

Authors : Patrick Kling, M. Tamer Ozsu, Khuzaima Daudjee,

University of Waterloo Cheriton School of Computer Science

**Forum : 37th International Conference** **on Very Large Data Bases, VLDB 2011**

**Download :** <http://vldb.org/pvldb/vol4/p1-kling.pdf>

Background

* Data Model: Directed Graph Representation
  + Assumptions of data model:
    - Ignores the distinction between XML elements.
    - Attributes by treating uniformly as nodes.
    - Original schema definition does not contain unspecified portions e.g ANY.
    - Extract the information capture by graph representation from DTD/ XML Schema.
    - Extracting schema information yields on schema graph which may be less restrictive than the original schema.
    - Schema graph is never used for the validation of documents.
* Query Model:
* Subset of XPath - Absolute location paths consisting of node tests with and without wildcards.
* Tree pattern representation of a query – Query Tree Pattern to express queries with multiple extraction points.
* Schema:
* Vertical Fragmentation Schema – Fragmenting the schema graph of the collection into connected sub-graphs. The schema graph may contain cycles and then the fragmentation schema be a DAG.
* Problems
* Distributed query evaluation.
* Localized, yielding local query plans.
* Pruning unnecessary fragments to eliminate unnecessary work.
* Focus
  + Similarly exploiting distribution in the context of XML
  + Generating distributed execution plans within the context of a vertically partitioned and distributed XML database system.
* Contributions
  + A technique for generating distributed execution plans within the context of a vertically partitioned and distribution XML database system.
  + Benefits and drawbacks of different types of execution plans
  + Optimization technique to prune irrelevant subtrees in a fragment by pushing crossing fragment joins into local query plans.
  + Compose the costs of local query plans while taking into account parallelism and optimization technique.

**Evaluation of the paper assessment**

* positive and negative sides
* unclear points etc. regarding contents, structure, language, figures etc

**Further comments**

* (typos, hints for improvements)