

# **SUPEX : A Schema-Guided Path Index for XML Data**

**Jing Wang**

**Institute of Computing Technology, CAS, China**

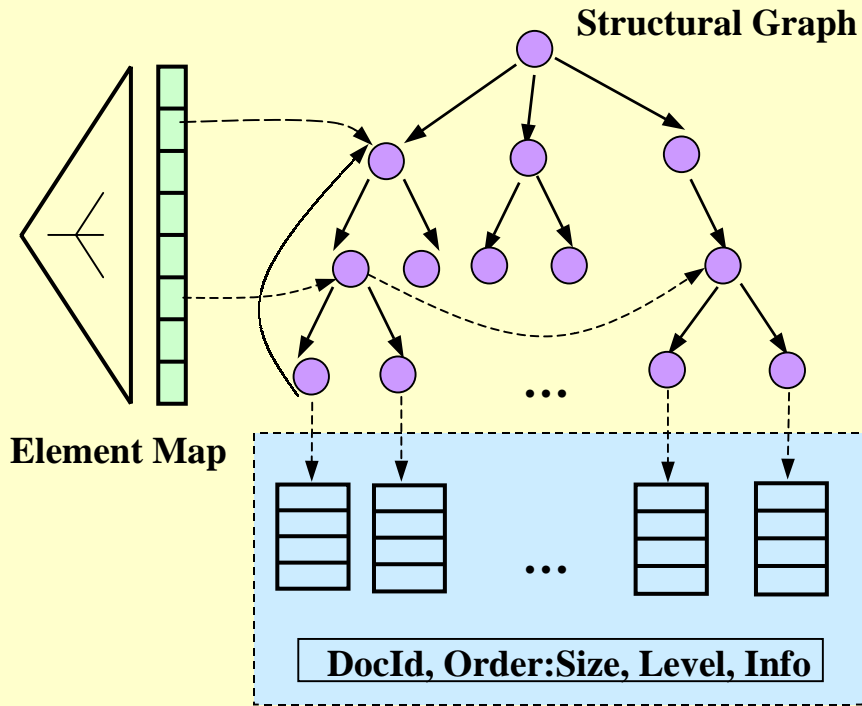
**Renmin University of China**

**Supervisor : Prof. Shan Wang, Prof. Xiaofeng Meng**

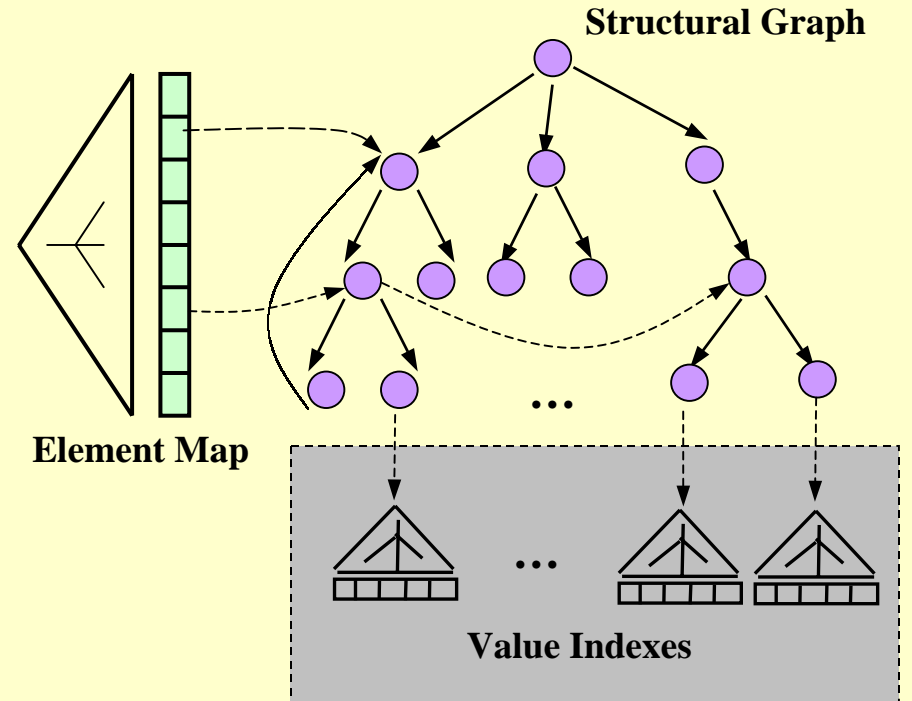
# Overview

- **Constructed based on DTD**
- **Integrating path index with value indexes**
- **Two main structures:**
  - **Structural Graph**  
the structure summary of XML data
  - **Element Map**  
fast entries to nodes in Structural Graph

# Architecture



**Path Index Structure**



**SUPLEX Structure With Value Indexes**

# The Procedure of Construction

## ■ Simplifying DTD

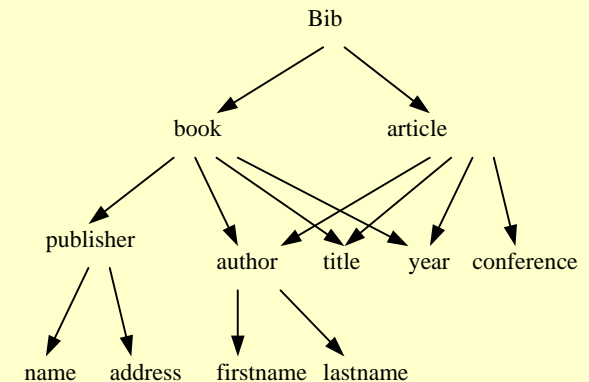
- $E1^* \rightarrow E1$
- $E1? \rightarrow E1$
- $E1+ \rightarrow E1$
- $E1 | E2 \rightarrow E1, E2$
- $\dots, E1, \dots, E1 \rightarrow \dots, E1, \dots$

## ■ Constructing structural graph

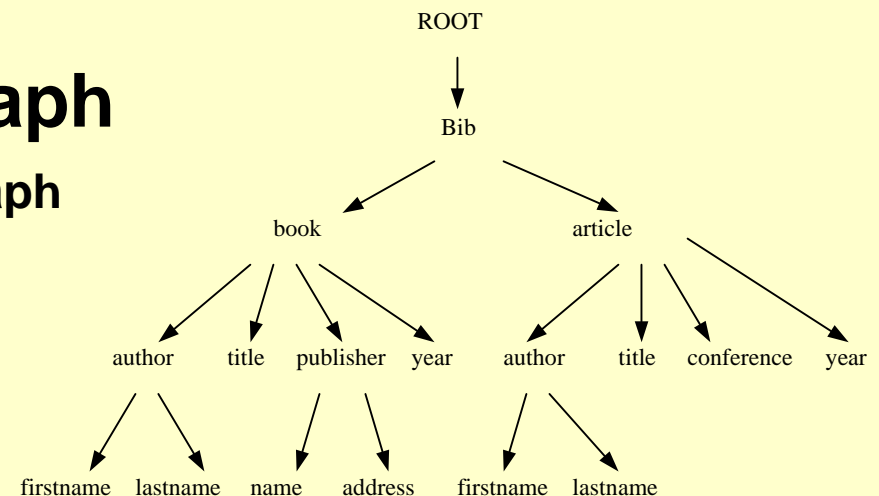
Generate structural graph from DTD graph

## ■ Loading data

During XML document loading



(a) DTD Graph

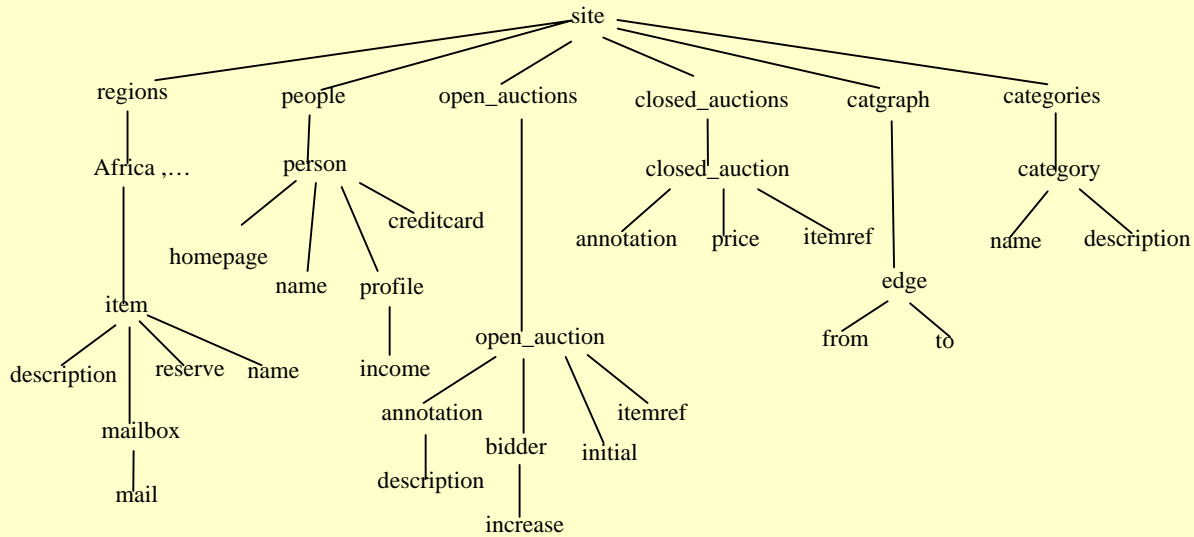


(b) Structural Graph

# Query Processing with SUPLEX

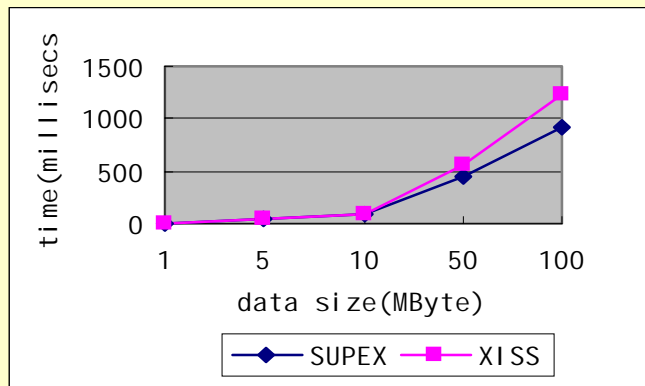
- **Two basic ways:**
  - Given a tag, all elements with this tag can be obtained by the lookup of EM.
  - Simple label paths from the root of document can be matched by traversal of SG starting from the root node.
- **Basic structural relationships:**
  - E1/E2
  - E1//E2
- **Partial label path:**
  - //E1/.../En
- **Query evaluation with value indexes**

# Preliminary Experiment Results

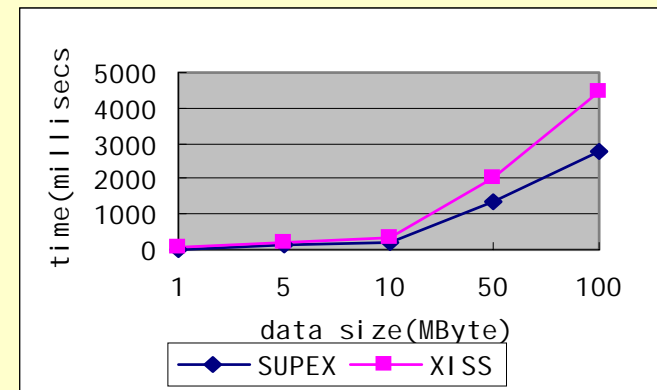


| Scaling Factor | Document Size(MB) | Element number |
|----------------|-------------------|----------------|
| 0.01           | 1.12              | 17132          |
| 0.05           | 5.6               | 83533          |
| 0.1            | 11.3              | 167865         |
| 0.5            | 56.2              | 832911         |
| 1.0            | 113               | 1666315        |

**Element hierarchy of XMark Document**



**open\_auction//description**



**description/text**

# Conclusion

- **Generated from XML schema information**
- **Concise path index**
- **Integrating path index with value indexes**
- **Supporting label path expressions**
- **Supporting the evaluation of value-based condition predicates**