OrientX: an Integrated, Schema-Based Native XML Database System

Meng Xiaofeng, Wang Xiaofeng, Xie Min, Zhang Xin, Zhou Junfeng

School of information, Renmin University of China

WISA2006

Introduction

- OrientX means:

  Original **R**UC **I**DK**E** **N**a**t**ive **X**ML Database

  – RUC: Renmin University of China

  – IDKE: Institute of Data and Knowledge Engineering

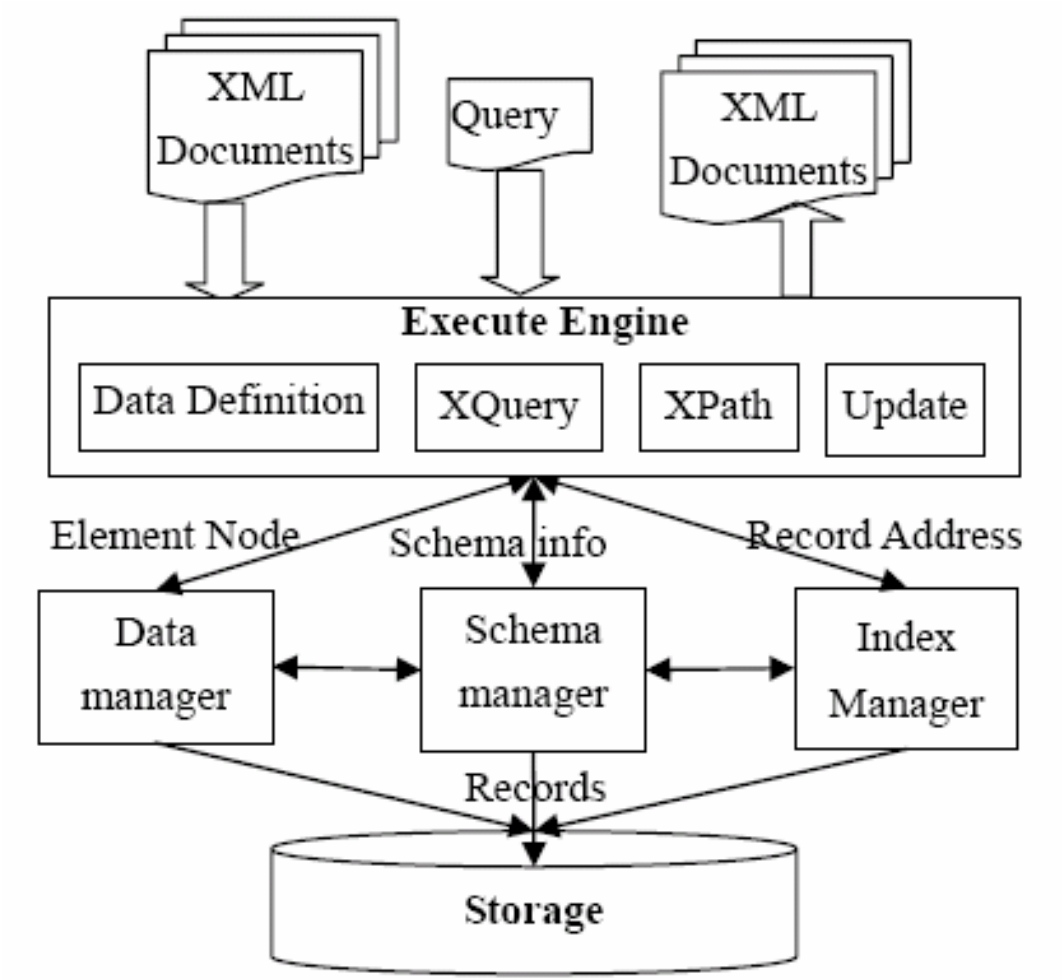
  – Native XML DataBase: Exposing a logical model of storing and retrieving XML documents.

  (non Native XML DataBase: for example, based on relation database)

Outline

- Architecture and Features
- Storage and data management
- Indexing Schema
- Query processing
- Conclusion and Future Work

Architecture



Features

- Full support to XML Schema
- Supporting XQuery1.0 and XPath2.0 Data Model
- Various native storage techniques
- Path index and value index
- Multi-Query Processing strategies based on native storage.

Outline

- Architecture and Features
- Storage and data management
- Indexing Schema
- Query processing
- Conclusion and Future Work

Different storage granularities

- Document:
  - do not decompose the document, build index on it to direct the structure.
  - Query complexity and efficiency are restricted by the power of index.
- Sub tree:
  - decompose the document into sub trees according to storage space partition.
  - Persistent the structure in the tree.
  - save space
- Node:
  - decompose the document into nodes sequence , each node corresponding to a type (element, attribute, …).
  - May use too many links to persistent relation between nodes

Storage Techniques in OrientX

	Element-based	SubTree-based	Document-based
Depth-first	**DEB**	**DSB**	DB
Broad-first	BEB	BSB	
Clustered	**CEB**	**CSB**	

One no through tree

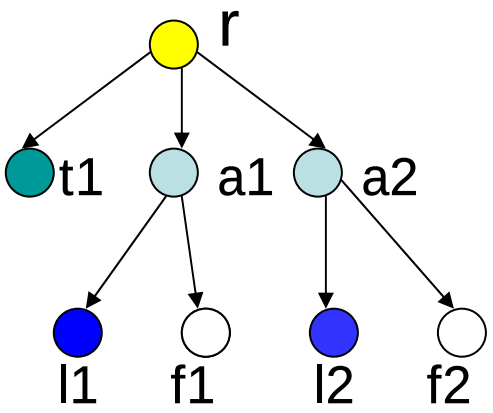
Like DE is a sub tree is page siz

One element all node with name will be

Akin to DSB, each record is a sub tree. But all sub trees with the same structure are clustered store.

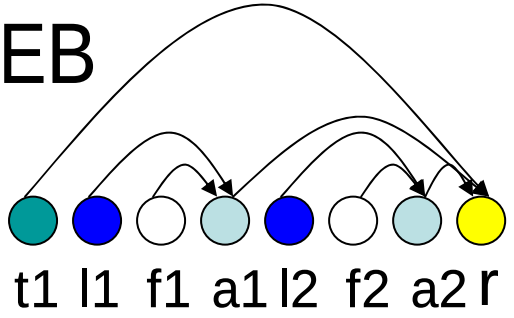
d in red

Example-- Element based

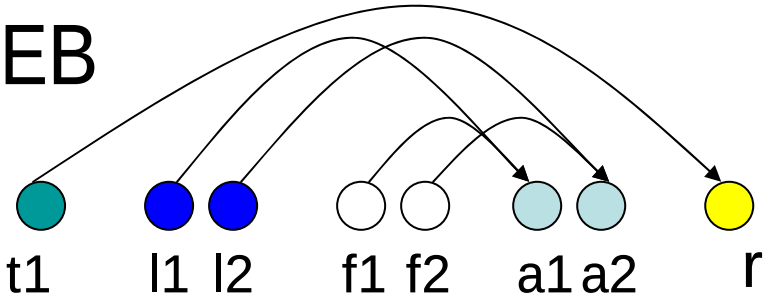


Source doc

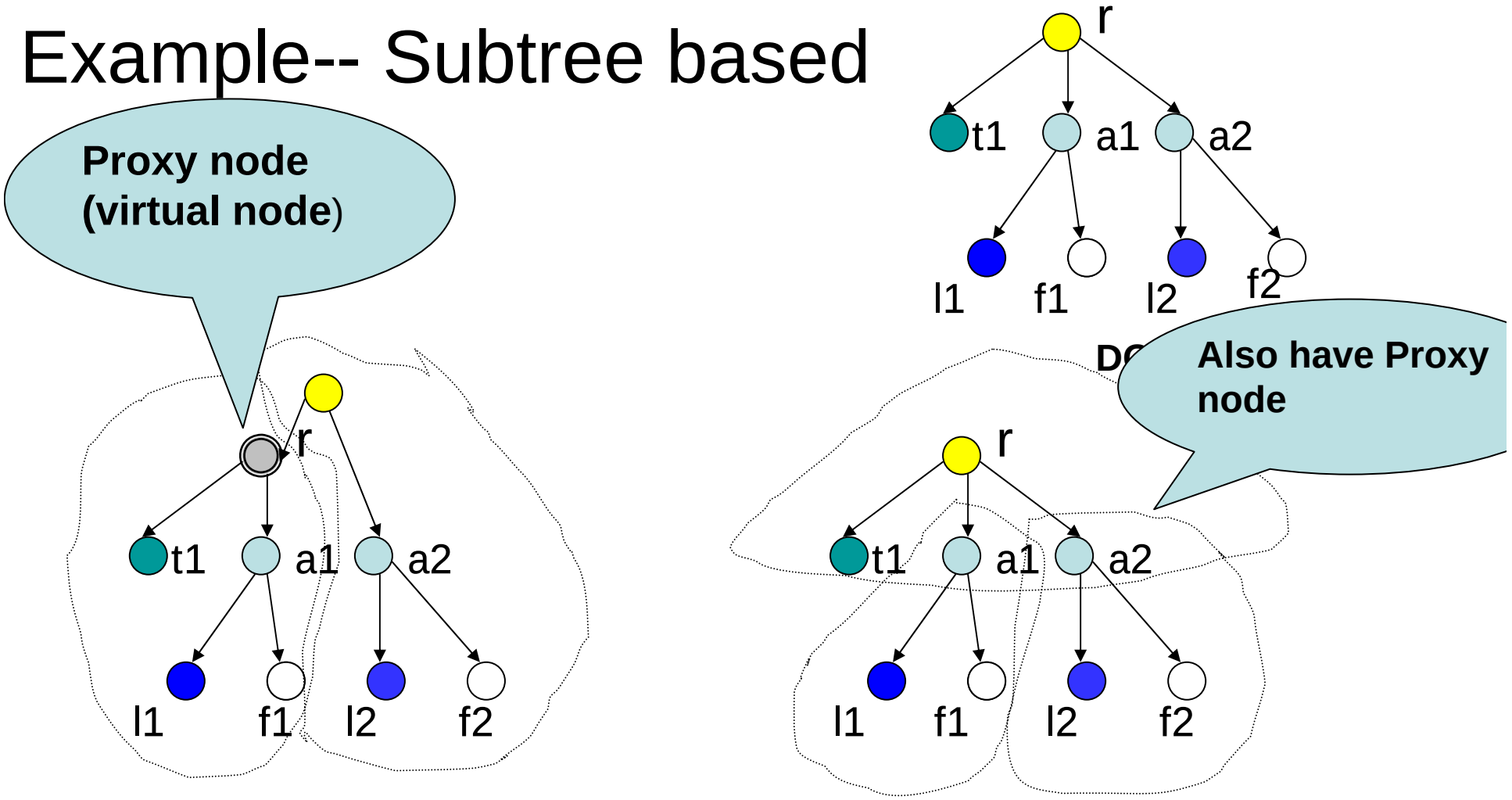
- DEB



- CEB



Example-- Subtree based



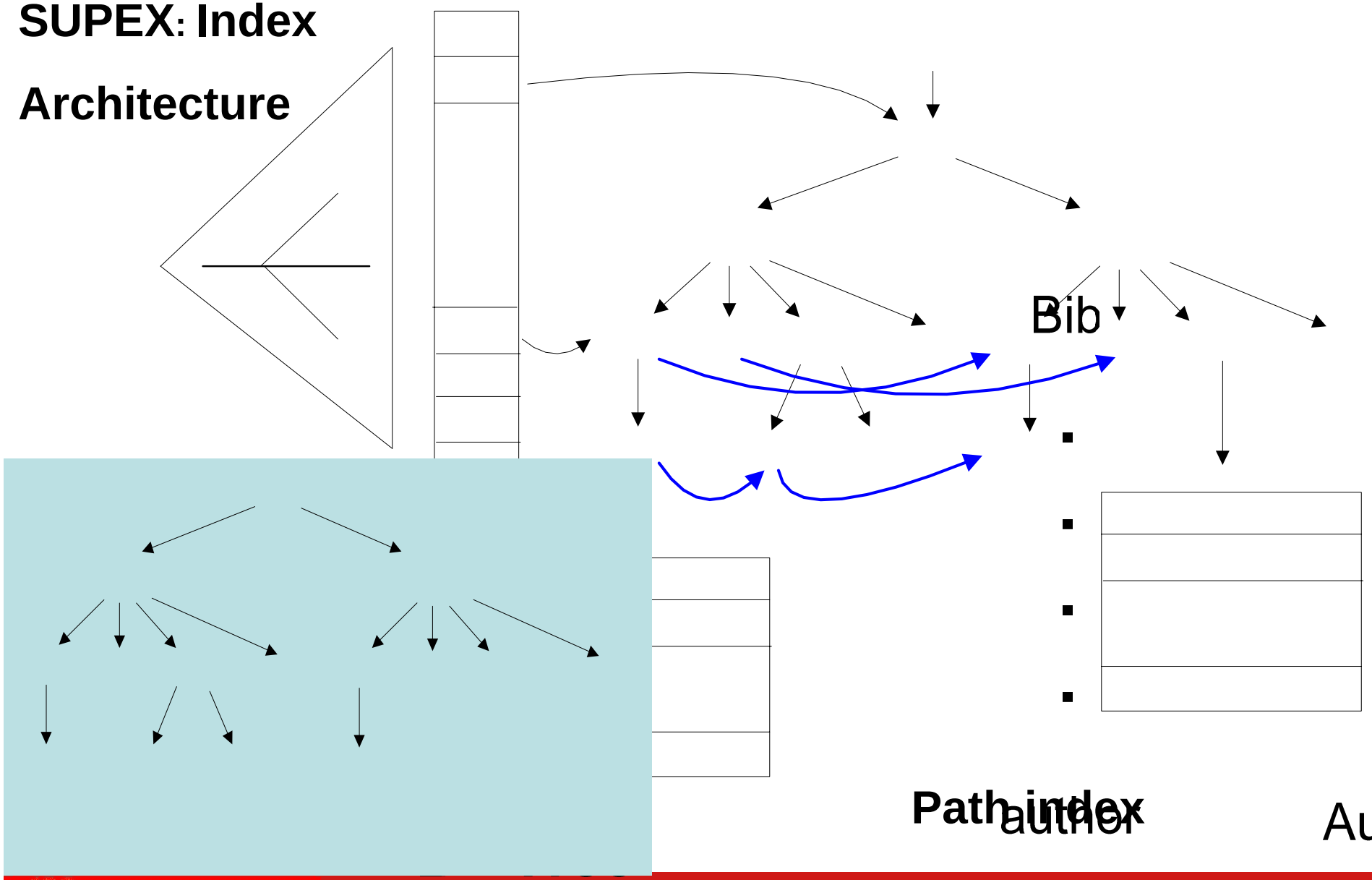
DSB(Depth-first sub-tree based)

CSB (clustered sub-tree based)

Outline

- Architecture and Features
- Storage and data management
- Indexing Schema
- Query processing
- Conclusion and Future Work

SUPEX: Index Architecture



Features of SUPEX

- Constructed based on DTD,Schema
- Integrating path index with value indexes
- Supporting Twig query efficiently
- Supporting label path expressions
  (bib//author)
- Supporting the evaluation of value-based condition predicates (//author[firstname = “jone”])

Outline

- Architecture and Features
- Storage and data management
- Indexing Schema
- Query processing
- Conclusion and Future Work

Query processing

- **Navigation strategy**
  - Supporting XPath2.0 and XQuery1.0
  - Combine continuous steps in one XPath into a single path.
  - Reform syntax tree into reduced execution plan.
  - Introducing the pipeline operator to XQuery process.

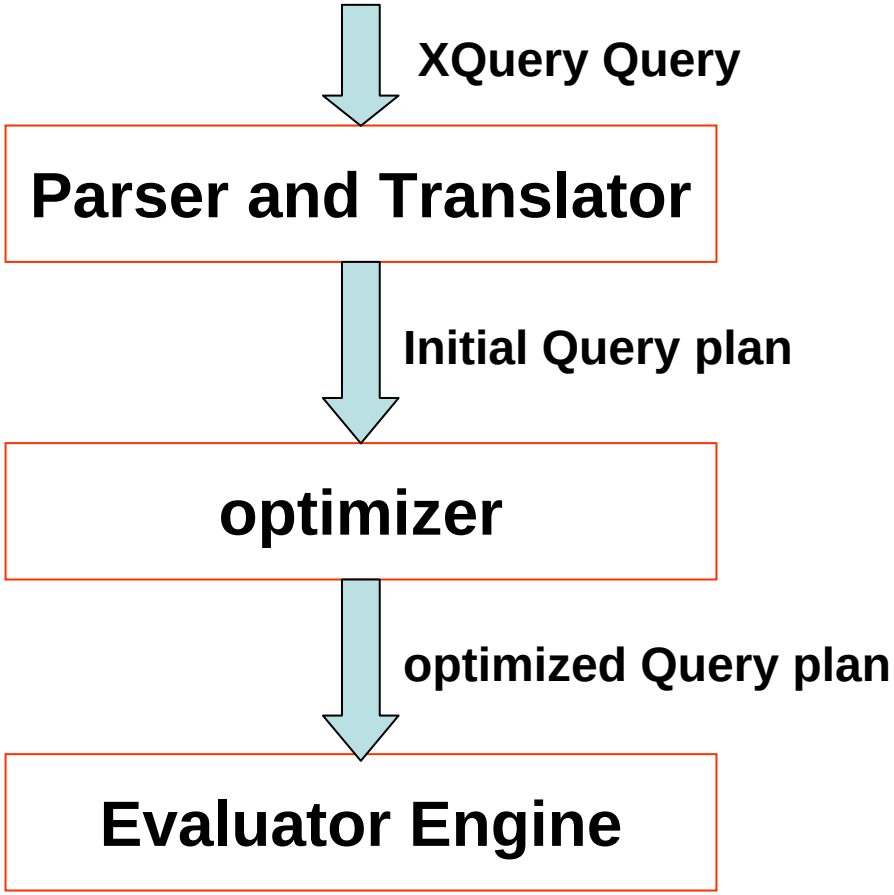
Operators in Navigation

```
<results>{
  let $doc := doc("docs/bib.xml")
  for $t in distinct-values($doc//book/title)
  let $p := $doc//book[title = $t]/price
  return
    <minprice title="{ $t }">
      <price>{ min($p) }</price>
    </minprice>
}</results>
```

Currently, Navigation Containing 13 operators:

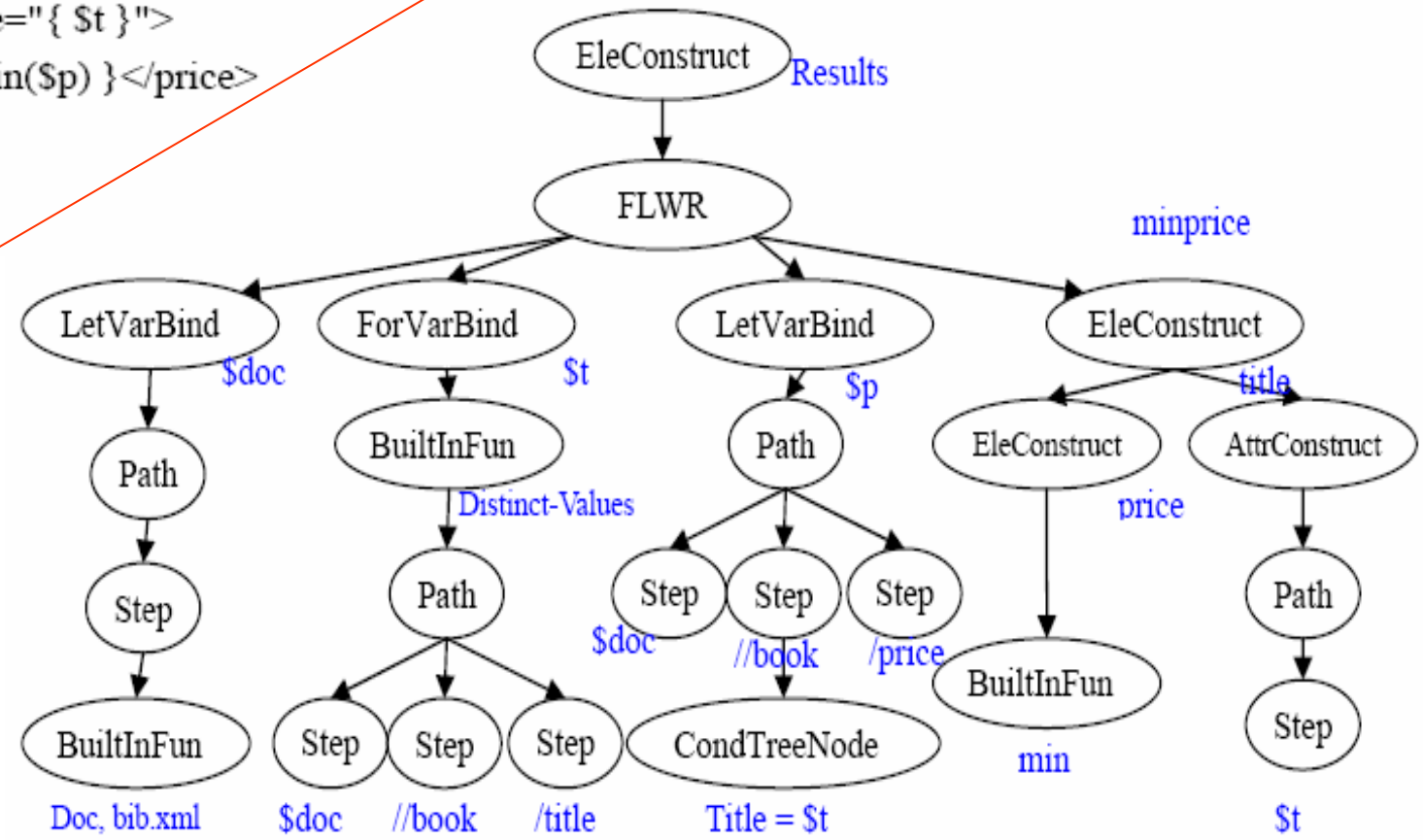
1. Step
2. CondTreeNode
3. Path
4. ForVarBind
5. LetVarBind
6. FLWR
7. EleConstructor
8. AttrConstructor
9. BuiltInFun
10. IfThenElse
11. Quanlify
12. SetOpt
13. SortBy

General Steps to process XQuery



```
<results>{
    let $doc := doc("docs/bib.xml")
    for $t in distinct-values($doc//book/title)
    let $p := $doc//book[title = $t]/price
    return
      <minprice title="{ $t }">
        <price>{ min($p) }</price>
      </minprice>
}</results>
```

The query plan



Outline

- Architecture and Features
- Storage and data management
- Indexing Schema
- Query processing
- Conclusion and Future Work

Conclusion and Future Work

- Conclusion:
  - OrientX is an integrated, schema-based native XML database system.
  - It implements storing and querying xml data.
- Future work:
  - XQuery optimization.
  - Xml Update and Other XQuery processing engine.

Thanks

Q&A☺

Welcome to our website http://idke.ruc.edu.cn

to obtain more information about OrientX