



X-Hive Corporation

[previous](#) | [next](#)

SEARCH METHODS FOR XML DOCUMENT COLLECTIONS

XML World, March, 27th 2001

Jeroen van Rotterdam, CTO X-Hive Corporation

www.x-hive.com



CONTENTS

- ✦ **XML datamodel**
- ✦ **Criteria for search methods**
- ✦ **A simple sample**
- ✦ **XML Query languages**
- ✦ **Selection criteria for Query languages**
- ✦ **Query language overview**
- ✦ **Conclusion**

www.x-hive.com

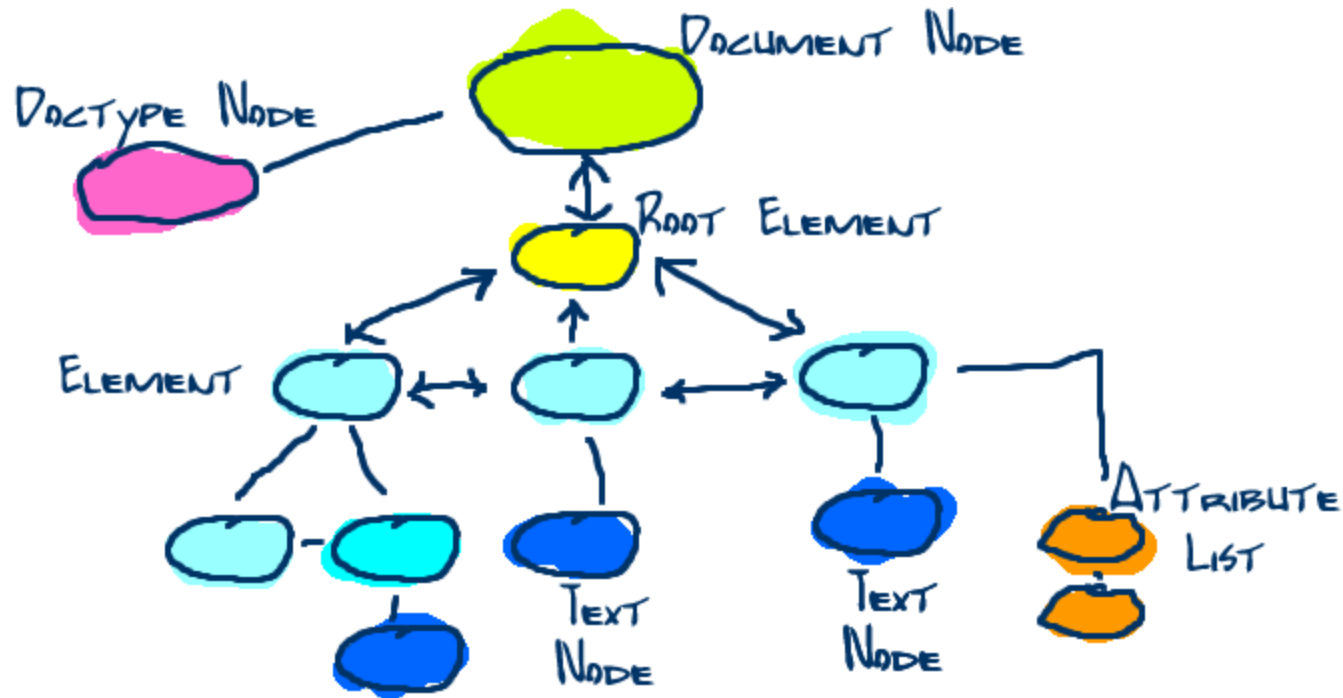


XML DATAMODEL

- **Tree oriented data model**
- **Document oriented versus table oriented**
- **Schema is a grammar not a table def**
- **Standard interface available (DOM)**
- **Several variants within different recommendations (e.g. XPath, XForms, XML Schema and DOM)**



DOM DATAMODEL



www.x-hive.com



TRADITIONAL SEARCH METHODS

- ✦ **SQL for relational model**
- ✦ **pattern matching (grep)**
- ✦ **inverted indexes for full text retrieval**



QUERY LANGUAGES CRITERIA IN GENERAL

- ✦ **efficient**
- ✦ **able to reflect user requests**
- ✦ **natural language aware**
- ✦ **easy to understand**
- ✦ **handling of simple as well as complex queries**
- ✦ **descriptive**
- ✦ **extensible**
- ✦ **standardized and adopted by the industry**

www.x-hive.com



XML QUERY LANGUAGES

- **should be context aware**
- **adapted to the XML data model**



CURRENT XML QUERY LANGUAGES / SEARCH METHODS

- ✦ **0-level search**
- ✦ **grep**
- ✦ **DOM Traversal/Treewalkers**
- ✦ **XPath**
- ✦ **XML-QL**
- ✦ **Quilt**
- ✦ **XQL**
- ✦ **XQuery**
- ✦ **XPointer**
- ✦ **etc.....**

www.x-hive.com



SAMPLE

```
<weather>
  <location city="Rotterdam">
    <measurement time="0810">
      <description>showers</description>
      <temp unit="celsius">4.0</temp>
      <wind>
        <direction>WNW</direction>
        <speed unit="km">11</speed>
      </wind>
    </measurement>
    <measurement time="1823">
      <description>partly cloudy</description>
      <temp unit="celsius">6.0</temp>
      <wind>
        <direction>W</direction>
        <speed unit="km">13</speed>
      </wind>
    </measurement>
  </location>
</weather>
```

www.x-hive.com



SEARCH REQUEST

What was the temperature in Rotterdam at 8.10 ?



O-LEVEL SEARCH METHOD

Too hard, don't want to know anymore.....

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
+++	---	+++	+++	---	+++	---	sometimes	+++	+++



GREP

- ✦ **pattern matching**
- ✦ **not content aware**

```
grep "\>.*</temp>" weather.xml
```

Nothing beats grep but...

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
---	--	---	---	---	---	---	---	---	---

www.x-hive.com



SQL

- ✦ not XML aware
- ✦ depends on the underlying proprietary datamodel
- ✦ does not reflect typical user requests in an XML environment
- ✦ largely adopted by the industry
- ✦ support for updates
- ✦ ISO standard

```
SELECT xml FROM mydocuments WHERE xml LIKE '%temp%'
```

or:

```
SELECT content FROM Element where Element.name=="...." AND .....
```

www.x-hive.com



SQL

Conclusion

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
---	---	---	+	---	++	---	+++	---	---



XQL

- ✦ not a standard
- ✦ not actively maintained
- ✦ XML-aware
- ✦ path extractions

Conclusion

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
+/--	+	---	--	+	---	---	---	+++	+++

www.x-hive.com



DOM TRAVERSAL

- ✦ **W3C Recommendation, DOM L2**
- ✦ **actively maintained (W3C DOM working group)**
- ✦ **not really a query language**
- ✦ **fully XML-aware**
- ✦ **developers only**
- ✦ **results are manipulable**



DOM TRAVERSAL

NodeFilter definition

```
public class MyFilter implements NodeFilter {  
  
    public short acceptNode (Node n) {  
  
        if( n.getNodeName().equals("temp") ){  
            Node parent = n.getParentNode();  
            if( parent != null ){  
                if( parent.getNodeName().equals("temp") ){  
                    Element e = (Element)parent;  
                    Attr    a = e.getAttribute("time");  
                    if( a.getNodeValue().equals("0810") ){  
                        return FILTER_ACCEPT;  
                    }  
                }  
            }  
        }  
        return FILTER_REJECT;  
    }  
}
```

www.x-hive.com



DOM TRAVERSAL

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
--	+	---	---	+++	---	++	+++	+++	+++



XPATH

- ✦ **W3C Recommendation**
- ✦ **actively maintained (W3C working group)**
- ✦ **XPath 2.0 currently under development**
- ✦ **fully XML-aware**
- ✦ **path extractions**
- ✦ **less descriptive but compact. (developers only ?)**
- ✦ **read only**
- ✦ **slightly different datamodel than the DOM**
 - ✦ **normalized view**
 - ✦ **parent of attributes**

www.x-hive.com



XPATH

Example:

```
//location[@city = 'Rotterdam']/measurement[@time = '0810']/temp
```



XPATH

Conclusion

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
+/-	+++	---	---	+++	---	+	+++	+++	+++



QUILT

- predecessor of XQuery
- it's there
- based on XQL extraction syntax
- variable binding mechanism from XML-QL
- currently not maintained
- no recommendation status

```
FOR    $a IN document("http://www.biblio.com/weather.xml")//location,  
      $b IN $a/measurement  
      $c IN $b/temp  
WHERE $a/@city = "Rotterdam" AND $b/@time = "0810"  
RETURN $b
```

www.x-hive.com



XQUERY

- according to the W3C query working group
- derived from Quilt
- based on XPath path extraction syntax
- variable binding mechanism from XML-QL
- major support within the working group
- 'descriptive' syntax



XQUERY

```
FOR $l in document("weather.xml")/location
    $m in $l/measurement
    $t in $m/temp
WHERE $l/@city = "Rotterdam" AND $m/@time = "0810"
RETURN
    <weather city="Rotterdam">
        <temperature> $t </temperature>
    </weather>
```

www.x-hive.com



XQUERY DRAWBACKS

- ✦ **no XML notation yet (feb 2001)**
- ✦ **not finished yet (feb 2001)**
- ✦ **not adopted yet**
- ✦ **no update queries**
- ✦ **no indexing (implementation issue)**
- ✦ **no sorting yet (feb 2001)**
- ✦ **another data model**



XQUERY

efficient	reflect req	natural lang	easy	simple vs. complex	descriptive	extensible	standard	context aware	XML datamodel
+/-	+++	---	+/-	+++	++	+	+++	+++	+++



XQUERY VS XML QUERY

- ✦ **XQuery is a subset of the XML Query requirements**
- ✦ **still a long list of open issues, including:**
 - ✦ **currently only a 'human readable' syntax, still to do: XML-based syntax**
 - ✦ **mapping of XQuery operators into algebra operators**
 - ✦ **alignment of XML Schema and XML Query/XQuery type systems + operators**



CONCLUSION

- ✦ **there's no single solution**
- ✦ **XML data model demands a different approach**
- ✦ **XPath, Treewalkers are currently available and useful**
- ✦ **XQuery is the road ahead**

However there is a need for:

- ✦ **update queries**
- ✦ **index support**
- ✦ **sorting facilities**
- ✦ **natural language awareness**

www.x-hive.com



X-Hive Corporation

[previous](#) | [next](#)

MORE INFO

- ✦ jeroen@x-hive.com
- ✦ www.x-hive.com/publications

www.x-hive.com

