

# Translating XSLT into XQuery

Albin Laga, Praveen Madiraju, Darrel A. Mazzari and Gowri Dara

Department of Mathematics, Statistics, and Computer Science

Marquette University

P.O. Box 1881, Milwaukee, WI 53201

{albin.laga, praveen.madiraju, darrel.mazzari, gowri.dara}@mu.edu

## ABSTRACT

XML (eXtensible Markup Language) transformations and queries are crucial operations for interpreting XML databases. XSLT (eXtensible Stylesheet Language Transformations) is a prominent XML technology for these operations, but XQuery (XML Query Language) can query a broad spectrum of XML information sources, including both databases and documents. In this paper, we present our approach to translate XSLT to XQuery. We illustrate our approach as a set of rules or templates which translates an XSLT document into XQuery. We also discuss grouping operations for both XSLT and XQuery.

**Keywords:** XSLT, XQuery, Xpath, XML Databases

## 1. INTRODUCTION

XML is becoming as a standard for the exchange of many data sources. The reason for the use of XML databases for database integration is, principally, together with W3C recommended query languages, such as XPath [3] and XQuery [4], and other techniques like XSLT [9] and SOAP (Simple Object Access Protocol), that XML helps to overcome most technical heterogeneities of diverse databases. In addition, XML databases can be transmitted and stored as text files with schema information transported using XML schemas or Document Type Definition files, unlike relational databases, though there may be speed degradation with very large text based databases. In order to extract information from XML databases, one solution would be to extend a knowledge modeling tool by implementing a set of new classes or functions in a language such as Java, though this can be a rather difficult and time consuming task. Another solution would be to use XSLT or XQuery for transforming XML documents. An advantage of this approach is that, even though an XSLT is written independently of any programming language, it can be *executed* by a program written in most modern programming languages [7].

XSLT and XQuery were developed in close collaboration, and therefore there is a high degree of functionality similarity of the two languages. They share many common concepts, such as the underlying data model.

They both include the whole of XPath as a sublanguage, which supports a number of data types and a respectable function library [5, 8].

The processing engine in XSLT automatically goes through the document tree and applies templates as it encounters nodes. XQuery, on the other hand, requires the programmer to direct the process. XSLT, in one sense, is like a “report generator” in which there exists an implicit processing cycle, and the programmer just sets up the actions that are desired to occur when certain conditions are met. XSLT is loosely typed as it is flexible in handling conversions between nodes, strings and numbers, for the most part, transparently. XQuery is a typed language which uses the types defined by XML Schema, and will complain when data types do not match with each other. XQuery implements the FLWR (For, Let, Where, & Result) expression. XQuery creates a list of binding tuples with the FOR clause which are taken from an ordered forest. Subsequent statements are executed once, including any additional binding tuples producing a cross product of these bindings. The construction of optimal bushy trees for computing the cross product of a set of relations is NP-hard. One of the problems the query optimizer has to deal with is the ordering of joins.

XSLT generates a current node list, with its “for-each” instruction, which is taken from a node set (XPath) in document order. In XQuery, like in a language such as “C”, the programmer is responsible for directing algorithms. XSLT uses the syntax of true XML, while XQuery is only XML-like. In Figure 1, we present a functional comparison between XSLT and XQuery. XSLT 2.0 [9] and XQuery 1.0 [4] target different user communities; the former is for transformation of one XML document into another, while the later targets querying of XML documents.

	XSLT		XQuery	
	Advantages	Disadvantages	Advantages	Disadvantages
Processing	Auto engine through document tree	Limited control	Programmer Directed	More effort, more possible errors
Data typing	Loosely typed	Parsing difficulties	Typed, more types possible	More effort, more possible errors

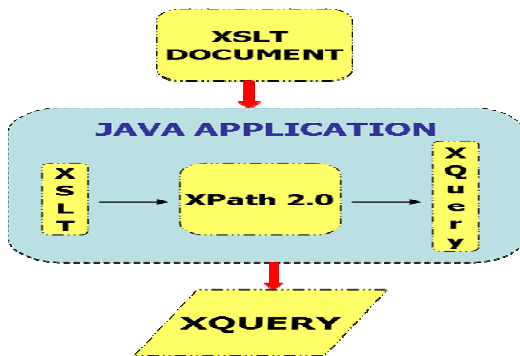
**Figure 1. A functional comparison of XSLT and XQuery**

As it stands now, XQuery 1.0 is not yet an official W3C recommendation, while XSLT 2.0 is an official recommendation of W3C. When W3C finalizes the XQuery 1.0 recommendation, we need to be able to translate some of the XSLT 2.0 (at least the ones which query) to XQueries. Hence, we need a tool to translate XSLT to XQuery. Also, it is predicted that most of the database systems will support XQuery, but may not implement XSLT. Furthermore, such a translation is needed for those XML database developers, who may understand XQuery but not XSLT[5].

The rest of the paper is organized as follows: In Section 2, we present our system architecture. We illustrate our approach for translating XSLT to XQuery with examples in Section 3. We then discuss grouping of data in both XSLT and XQuery in detail in Section 4. Finally, in Section 5 we describe related work and present our conclusions.

## 2. SYSTEM ARCHITECTURE

The system architecture is shown in Figure 2. The application, written in Java, takes an XSLT 2.0 [9] template as input and outputs an XQuery 1.0 document [4]. As both XSLT 2.0 and XQuery 1.0 have XPath 2.0 as their subset, wherever equivalent methods exist in both of them, the application will utilize XPath 2.0 for transformations directly from XSLT 2.0 to XQuery 1.0. In the case where an equivalent method does not exist, functions that mimic the XSLT 2.0 methods will be generated for XQuery 1.0.



**Figure 2. Translating XSLT into XQuery architecture**

XPath 2.0 is an expression language that allows processing of values conforming to the data model defined in the XQuery/XPath Data Model (XDM). The data model provides the XML documents in the form of a tree representation, and also as atomic values such as integers, strings, booleans, and sequences that may contain both references to nodes in an XML document or atomic values. The result of an XPath expression may be a selection of nodes from the input documents, or an atomic value, or, more generally, any sequence allowed by the data model. XPath navigates through the hierarchical structure of an XML document and hence the name [3].

## 3. OUR APPROACH

Throughout the paper, we illustrate our approach of translation using a sample XML document shown in Figure 3.

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!-- Edited with XML Spy v4.2 -->
<catalog>
  <cd>
    <title>Empire Burlesque</title>
    <artist>Bob Dylan</artist>
    <country>USA</country>
    <company>Columbia</company>
    <price>10.90</price>
    <year>1985</year>
  </cd>
  <cd>
    <title>Hide your heart</title>
    <artist>Bonnie Tyler</artist>
    <country>UK</country>
    <company>CBS Records</company>
    <price>9.90</price>
    <year>1988</year>
  </cd>
</catalog>
```

**Figure 3. A fragment of cdcatalog.xml; a full version of the document is available from [15]**

XPath is a rich language with many built-in functions included in it. XSLT and XQuery are closely related because the underlying path language for both XSLT and XQuery is XPath. Hence, the translation is much easier as they share the same built-in functions [5].

### 3.1 Translation Rules

Here, we show the translations of each XSLT construct into equivalent XQuery via the examples on the cdcatalog.xml (refer to Figure 3). We have tested for the correctness of all our XSLTs and XQueries using Saxon 8.7 [8]. Due to space limitations, we cannot cover the entire language details of XSLT transformation into XQuery. We will cover the most important and also challenging translations (for example, grouping).

**Rule 1:** XQuery and XSLT support declaration of parameters and variables. They can be declared and

invoked in both languages. However, XQuery variables are more restricted because they have strict type.

XSLT	XQuery
<pre>&lt;xsl:param name="cd"&gt;Cd collection&lt;/xsl:param&gt; &lt;xsl:variable name="cde"&gt;My Cd collection &lt;/xsl:variable&gt;</pre>	<pre>declare variable \$c as xs:string := "Cd collection";</pre>

**Rule 2:** The XSLT function `<xsl:value-of>` is translated into a XQuery expression `for $x in ... (collection of data[1])`. Since “for” clause returns a collection, we narrow the retrieved values to the very first one by explicitly asking for the first value of that collection.

XSLT	XQuery
<pre>&lt;xsl:value-of select="catalog/cd/title" /&gt;</pre>	<pre>for \$x in ... (collection of data[1]) /&gt;</pre>

**Rule 3:** The XSLT function `<xsl:for-each>` is translated into XQuery expression `for $x in ...` as shown below:

XSLT	XQuery
<pre>&lt;xsl:for-each select="node"&gt; &lt;!--function Body --&gt; &lt;/xsl:for-each&gt;</pre>	<pre>for \$x in ... (collection of data)</pre>

**Rule 4:** Filtration is very straight forward in both languages and can be specified in the square brackets in an XPath expression. Alternatively, in XQuery we can simply use a “where” clause.

XSLT	XQuery
<pre>XPath/expression/[where condition]/node</pre>	<pre>XPath/expression/[where condition]/node Or where variable = “something”</pre>

**Rule 5:** Sorting is implemented in XSLT by using the function `<xsl:sort/>` and this must be placed as the very first one in the statement designated to retrieve a set of values. In XQuery we use “order by” clause.

XSLT	XQuery
<pre>&lt;xsl:sort select="element"/&gt;</pre>	<pre>Order by variable/XPath/expression</pre>

**Rule 6:** The XSLT function `<xsl:if>` is translated into XQuery by `if () then ... else()` function while the `else()` statement must be included with empty parentheses.

XSLT	XQuery
<pre>&lt;xsl:if test="conditional expression"&gt; &lt;!--function Body --&gt; &lt;/xsl:if&gt;</pre>	<pre>If( conditional expression ) then &lt;!--function Body --&gt; else()</pre>

**Rule 7:** The XSLT function `<xsl:choose>` is translated into XQuery by the following function `if() then <!--function Body --> else() if() then <!--function Body --> else() <!--function Body -->`. We can have many `<xsl:when>` elements but only one `<xsl:otherwise>` element in a `<xsl:choose>` function. Similarly, in XQuery we can have many `if() then <!--function Body -->` elements but only one `else() <!--function Body -->` element in a `if() then` function.

XSLT	XQuery
<pre>&lt;xsl:choose&gt; &lt;xsl:when test=" conditional expression "&gt; &lt;!--function Body --&gt; &lt;/xsl:when&gt; &lt;xsl:when test=" conditional expression "&gt; &lt;!--function Body --&gt; &lt;/xsl:when&gt; &lt;xsl:otherwise&gt; &lt;!--function Body --&gt; &lt;/xsl:otherwise&gt; &lt;/xsl:choose&gt;</pre>	<pre>if(\$x/price &gt; 10) then &lt;!--function Body --&gt; if (\$x/price &gt; 9) then &lt;!--function Body --&gt; else() &lt;!--function Body --&gt;</pre>

**Rule 8:** Templates in XSLT are translated into user defined functions in XQuery. This topic is covered in depth in [5].

XSLT	XQuery
<pre>&lt;xsl:template match="node"&gt; &lt;!--Template Body --&gt; &lt;/xsl:template&gt;</pre>	<pre>declare function local:cd-info(\$cd as element()? ) as element()? { &lt;!-- Function Body --&gt; }</pre>

Grouping of data can be achieved in both XSLT and XQuery. We illustrate these operations in the next section. Thusly the application presents a general method for translating the highly declarative rule-base approach of XSLT into the purely functional XQuery approach.

#### 4. GROUPING IN XSLT AND XQUERY

Grouping of data is a sophisticated operation. Here, we show grouping in XSLT and then an equivalent way of implementing it in XQuery.

## 4.1 Grouping in XSLT 2.0

Here, we explain the process of grouping data from an XML document in detail using XSLT. XSLT 1.0 has a roundabout way of implementation of grouping [13]. Fortunately, in XSLT 2.0, grouping has been simplified by introducing `xsl:for-each-group` element. In Figure 4, we show an example XSLT which selects artist and title information grouping by each country.

The return value of `xsl:for-each-group` is a *sequence of groups*. This feature makes this function very efficient since it returns a sequence of groups at one time. This is one of the main reasons why grouping in XSLT 2.0 works faster than grouping in XQuery 1.0 (see performance results in [10]).

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="2.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:template match="catalog">
    <html>
      <body>
        <h2>My CD Collection</h2>
        <ul>
          <xsl:for-each-group select="cd" group-by="country">
            <b> <xsl:value-of select="country"/> </b>
            <xsl:for-each select="current-group()">
              <p> <xsl:apply-templates select="title"/>
                <xsl:apply-templates select="artist"/> </p>
            </xsl:for-each>
            <p>CD# = <xsl:value-of select="count(current-
group()/title)"/>
            </p>
          </xsl:for-each-group>
        </ul>
      </body>
    </html>
  </xsl:template>
  <xsl:template match="cd">
  </xsl:template>
  <xsl:template match="title">
    Title: <span style="color:#ff0000">
  <xsl:value-of select="."/>
  </span>
  <br/> </xsl:template>
  <xsl:template match="artist">
    Artist: <span style="color:#00ff00">
  <xsl:value-of select="."/>
  </span>
  <br/>
  </xsl:template>
</xsl:stylesheet>
```

Figure 4: An example XSLT illustrating grouping operation

## 4.2 Grouping in XQuery 1.0

There is no dedicated function for grouping in XQuery 1.0. In order to group data in XQuery 1.0, we need to follow these two steps: (i) create the list of distinct values (grouping keys) by which we want to group our data, and save the result in a sequence variable, and (ii) in the *return* statement, place the sequence variable from the earlier step and then, use the current value of this variable as an argument for the *where* clause of the next loop statement. This statement returns the group element that satisfies the condition of *where* clause.

If we want to use an aggregation function, XQuery has to construct the group again. The example in Figure 6 is an equivalent transformed XQuery 1.0 document for the XSLT shown in Figure 5.

```
(: declare functions :)
declare function local:cd-info($cd as element()?) as
element()?
{
  let $t := $cd/title
  let $a := $cd/artist
  return
  <p>Title:<span
style="color:#ff0000">{data($t)}</span>
  <br/>
  Artist:<span
style="color:#00ff00">{data($a)}</span>
  </p>
};
<html>
<body>
<h2>Cd Collection</h2>
{
  for $c in distinct-values(doc("cdcatalog.xml")
/catalog/cd/country)
  return
  <p>
    { $c }
    {
      for $x in doc("cdcatalog.xml")/catalog/cd
      where $x/country = $c
      return <li>{ local:cd-info($x)}</li>
    }
    CD# = { count(doc("cdcatalog.xml")/catalog/cd
[country = $c]) }
  <p/>
  </p>
}
</body>
</html>
```

Figure 5: An equivalent transformed XQuery for the XSLT in Figure 4

In Figure 5, the second loop of the *return* statement has to re evaluate as many times as the number of elements in variable \$c. Every time the *return* statement loops, it creates a new group, and this process increases the execution time of XQuery and makes it less efficient than XSLT 2.0. We can say that XSLT 2.0 creates the **sequence of groups** only once, while XQuery creates the group as many times as the *return* statement loops. This is one of the main reasons why XQuery 1.0 is slower in comparison to XSLT 2.0.

## 5. RELATED WORK AND CONCLUSIONS

**Related Work:** There are many software packages available in the market such as [8, 16], which support both XSLT and XQuery executions. None of the commercial software packages support automatic translations. Fokoue et al. [5] present a detailed overview of compiling an XSLT 2.0 document into XQuery 1.0 document. The theoretical treatment especially is clear. However, the authors do not address grouping and sorting of XSLT's into XQueries.

Bezivin et al. [2] describe an approach to transform XSLT into XQuery using Atlas Transformation Language (ATL) within the Model Driven Architecture (MDA). They present more emphasis on the language details and do not go into details about the transformation approach with examples such as grouping and others. However, we present our approach in a simple form with clear examples all throughout the paper. The authors in [11] present translation from XSLT to SQL queries. Moerkotte (2002) [12] describes an implementation to transform XSLT to database algebra.

**Conclusions:** We have presented a general framework and an approach to translate XSLT 2.0 to XQuery 1.0. We have illustrated our approach with many examples. As XQuery becomes more popular and attracts querying for XML documents, our tool will help XML database developers translate current XSLT documents to XQuery. Due to space limitations, we are not able to show the performance evaluations. However, interested readers can refer to the extended version of this paper [10]. This study gives some open problems such as XQuery optimizations, and improving grouping functions in XQuery. We also intend to stress our translator to all sets of XSLT use cases. We will also keep our eye open for new changes in XQuery as it gets its official recommendation from W3C. Accordingly, we will modify our translator.

## 6. REFERENCES

- [1]. Apache Xalan. <http://xml.apache.org/xalan-j/>, last accessed on March 28, 2006
- [2]. Bézivin, J., Dupé, G., Jouault, F., Pitette, G. and Rougui, J.E. First experiments with the ATL model transformation language: Transforming XSLT into XQuery, OOPSLA 2003 Workshop, Anaheim, CA.
- [3]. Berglund A., Boag S, Chamberlin D, Fern´andez M. F., Kay M, Robie J, and Sim´eon J. (2004) XML path language (XPath) 2.0. W3c working draft, World Wide Web Consortium, July 2004.
- [4]. Boag, S., Chamberlin, D., Fernandez, M.F., Florescu, D., Robie, J., and Simeon, J. XQuery 1.0: An XML query language. W3c working draft, World Wide Web Consortium, July 2004. <http://www.w3.org/TR/2004/WD-xquery-20040723>
- [5]. Fokoue, A., Rose, K., Siméon, J., and Villard, L. Compiling XSLT 2.0 into XQuery 1.0. "In Proceedings of the Fourteenth International World Wide Web Conference, Japan, May 05, pp. 682-691.
- [6]. Jain, S., Mahajan, R., and Suciu, D. 2002. Translating XSLT programs to Efficient SQL queries. In Proceedings of the 11th international Conference on World Wide Web , Hawaii, USA, May 2002
- [7]. Jovanovic, J., and Gasevic, D. Achieving knowledge interoperability: An XML/XSLT approach. Expert Systems with Applications, Volume 29, Issue 3, October 2005, Pages 535-553.
- [8]. Kay, M. (2005). Comparing XSLT and XQuery. Proceedings of the XTech 2005 Conference on XML, the Web and beyond ; software available from : <http://saxon.sourceforge.net/>, last accessed on March 29, 2006
- [9]. Kay, M. XSL transformations (XSLT) version 2.0. W3c working draft, World Wide Web Consortium, November 2003. <http://www.w3.org/TR/2003/WD-xslt20-20031112>.
- [10]. Laga, A., Madiraju, P., Mazzari, D.A., and Dara, G. Translating XSLT into XQuery. Unpublished manuscript available from : <http://www.mscs.mu.edu/~praveen/Research/XSLT2XQ/>
- [11]. Liu, J., and Vincent, M. Query Translation from XSLT to SQL. Seventh International Database Engineering and Applications Symposium , 2003.
- [12]. Moerkotte, G. Incorporating XSL processing into database engines. In VLDB , September, 2002, Hong Kong, China, pp. 107–118.
- [13]. Tennison, J. Grouping Using the Muenchian Method. <http://www.jenitennison.com/xslt/grouping/muenchian.xml>, last accessed December 12, 2005
- [14]. The Sun Multi-Schema XML Validator(MSV). <https://msv.dev.java.net/>, last accessed on March 28, 2006
- [15]. W3 School, XML Application. [http://www.w3schools.com/xml/cd\\_catalog.xml](http://www.w3schools.com/xml/cd_catalog.xml), last accessed on December 12, 2005.
- [16]. XML Spy. XML Spy 2005. [http://www.altova.com/products\\_ide.html](http://www.altova.com/products_ide.html), last accessed on March 28, 2006