Automatic Generation of SQL/XML Views

Vânia Maria Ponte Vidal¹, Marco Antonio Casanova², Fernando Cordeiro Lemos¹

¹Department of Computing – UFC
Fortaleza – CE – Brazil

²Department of Informatics – PUC-Rio
Rio de Janeiro – RJ – Brazil

vvidal@lia.ufc.br, casanova@inf.puc-rio.br, fernandocl@lia.ufc.br

Abstract. This paper proposes an approach to generate XML views of relational data, using SQL/XML. The paper first specify the conditions for a set of correspondence assertions to fully specify the view in terms of the relational schema and, if so, we show that the mappings defined by the view correspondence assertions can be expressed as SQL/XML view definition. This paper focuses on an algorithm that automatically generates the SQL/XML query from the view correspondence assertions.

1. Introduction

XML has emerged as the standard information exchange format for Internet-based business applications. However, since most business data is currently stored in relational database systems, the problem of publishing relational data in XML format has special significance. A general and flexible way to publish relational data in XML format is to create XML views of the underlying relational data.

The exported view may be either virtual or materialized. Materialized views improve query performance and data availability, but they must be updated to reflect changes to the base source [12]. In the case of virtual views, the data still persists in relational databases, while applications may access the data in XML format through the XML view [1]. Exporting virtual XML views of relational data raises the problems of defining the XML view and evaluating an XML query posed over the view. The XML query is translated into SQL by composing it with the view definition.

The publication of relational data through a virtual XML view has been addressed, for example, in XPeranto [3] and SilkRoute [6]. In both works, the XML view is defined as an XQuery over the canonical XML view that represents the database tables and their attributes. This query specifies the view schema and the mapping knowledge, describing how the schema is related to the canonical view. The evaluation of an XML query over the view is performed using a middleware on top of relational database. The middleware translates the XML query into equivalent SQL queries. Then, the SQL results are tagged to produce the resulting XML document. In these systems, efficient query processing is not guarantee.

In the DB2 XMLExtender [2] and in SQL Server [11], the mapping knowledge is stored within annotated schemas [11]. In both cases, the mapping definition is very complex. Moreover, SQL Server provides the FOR XML clause to provide modes to transform query results into XML. The mapping knowledge is defined at access time and not stored in any way, which violates the mapping transparency.
With the introduction of the XML datatype [1] and the SQL/XML standard [4] as part of SQL:2003 [4], users may resort to the SQL/XML publishing functions to create virtual XML views over base relational schemas. Oracle [1] was the first DBMS to support, with its XML DB module [1], the creation of XML Views as SQL/XML queries over the relational data. The advantages of this approach rely on the use of a standard to publish relational data and on the capacity to process the SQL/XML publish functions within the SQL statements, which represents a gain in performance [9]. Thus, XML Query rewrite can be performed inside the DBMS [8], as opposed to non-integrate mid-tier solution.

However, creating SQL/XML view definitions demands advanced knowledge of SQL/XML and is time consuming. Moreover, users will have to redefine the XML view whenever the base relational schema changes. Therefore, tools that facilitate the task of XML view creation and the maintenance should be developed.

We propose in this paper an approach where the SQL/XML view definition is derived from view correspondence assertions, which specify relationships between the view schema and the relational schema. In the case of materialized views, as we shown in [12], all rules required to maintain the view can be automatic generated based on the view correspondence assertions.

This paper has three major contributions. First, we propose the use of correspondence assertions [10][12] for specifying the mapping between an XML view schema and a base relational schema. We formally specify the conditions under which a set of correspondence assertions fully specifies the XML view in terms of the relational source and, if so, we show that the mappings defined by the view correspondence assertions can be expressed as an SQL/XML query view definition. Second, we propose an algorithm that, based on the view correspondence assertions, generates the SQL/XML query that constructs the XML view elements from the relational tuples. Third, we propose the XMLView-By-Assertions (XVBA) tool that facilitates the task of XML view creation and maintenance. We note that the mapping formalisms used by other schema mapping tools are either ambiguous [7] or require the user to declare complex logical mapping [14].

This article is organized as follows. Section 2 discusses XML Views and the SQL/XML standard. Section 3 presents our mapping formalism. Section 4 discusses how to specify XML view using correspondence assertions. Section 5 presents the algorithm that automatically generates the SQL/XML view definition from the correspondence assertions. Finally, Section 6 presents the conclusions.

2. XML Views

With the introduction of the XML datatype and the SQL/XML standard, users may create a view of XML type instances over relational tables using SQL/XML publishing functions [1], such as XMLElement(), XMLConcat(), etc.

Consider, for example, the relational schema ORDERS_DB and the XML type PurchaseOrder_Type, whose graphical representations are shown in Figure 1 and 2 respectively. To generate instances of PurchaseOrder_Type from ORDERS_DB, we create the SQL/XML view PurchaseOrder_XML shown in Figure 3. As illustrated in Figures 4 and 5, for each tuple in table ORDERS_REL, the XML view uses the SQL/XML standard publishing functions to construct an instance of the XMLType PurchaseOrder_Type.
CREATE OR REPLACE VIEW PurchaseOrder_XML AS
SELECT XMLELEMENT("PurchaseOrder",
    XMLATTRIBUTES(O.ORDER_NO AS "ID")
) AS "ID",
    XMLFOREST(O.ORDER_DATE AS "OrderDate")
) AS "OrderDate",
    XMLFOREST(C.CUST_NAME AS "Name")
) AS "Name",
    XMLFOREST(C.STREET AS "Street")
) AS "Street",
    XMLFOREST(C.CITY AS "City")
) AS "City",
    XMLFOREST(C.STATE AS "State")
) AS "State",
    XMLFOREST(C.ZIP AS "ZIP")
) AS "ZIP",
    XMLFOREST(C.PHONE1 AS "Phone")
) AS "Phone",
    XMLFOREST(C.PHONE2 AS "Phone")
) AS "Phone",
    XMLFOREST(C.PHONE3 AS "Phone")
) AS "Phone"
FROM CUSTOMERS_REL C
WHERE C.CUST_NO = O.CUST_NO,
    XMLFOREST(L.ITEM_NO AS "ItemNo")
) AS "ItemNo",
    XMLFOREST(D.NAME AS "Name")
) AS "Name",
    XMLFOREST(D.PRICE AS "Price")
) AS "Price",
    XMLFOREST(D.TAX_RATE AS "TaxRate")
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate"
) AS "TaxRate";

WHERE O.ORDER_NO = P.ORDER_NO,
    XMLFOREST(L.QUANTITY AS "Quantity")
) AS "Quantity",
    XMLFOREST(L.DISCOUNT AS "Discount")
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount"
) AS "Discount";
CUSTOMERS_REL

<table>
<thead>
<tr>
<th>CUST_NO</th>
<th>CUST_NAME</th>
<th>STREET</th>
<th>CITY</th>
<th>STATE</th>
<th>ZIP</th>
<th>PHONE1</th>
<th>PHONE2</th>
<th>PHONE3</th>
</tr>
</thead>
<tbody>
<tr>
<td>193</td>
<td>Bryan Huston</td>
<td>8 Automation Ln</td>
<td>Albany</td>
<td>NY</td>
<td>12205</td>
<td>+91 11 012 4813</td>
<td>+91 11 083 4813</td>
<td>+91 33 012 4827</td>
</tr>
<tr>
<td>195</td>
<td>Cary Stockwell</td>
<td>400 E Joppa Rd</td>
<td>Baltimore</td>
<td>MD</td>
<td>21286</td>
<td>+91 11 012 4835</td>
<td>NULL</td>
<td>NULL</td>
</tr>
</tbody>
</table>

PRODUCTS_REL

<table>
<thead>
<tr>
<th>PROD_NO</th>
<th>NAME</th>
<th>PRICE</th>
<th>TAX_RATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2638</td>
<td>HD 10GB</td>
<td>125.50</td>
<td>0.01</td>
</tr>
<tr>
<td>1721</td>
<td>PC Bag - L/S</td>
<td>256.28</td>
<td>0.005</td>
</tr>
<tr>
<td>1761</td>
<td>Mouse +WP/CL</td>
<td>32.89</td>
<td>0.0</td>
</tr>
</tbody>
</table>

LINE_ITEMS_REL

<table>
<thead>
<tr>
<th>ORDER_NO</th>
<th>ITEM_NO</th>
<th>PROD_NO</th>
<th>QUANTITY</th>
<th>DISCOUNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>405</td>
<td>1</td>
<td>2638</td>
<td>35</td>
<td>0.07</td>
</tr>
<tr>
<td>407</td>
<td>1</td>
<td>1721</td>
<td>15</td>
<td>0.05</td>
</tr>
<tr>
<td>408</td>
<td>1</td>
<td>1721</td>
<td>30</td>
<td>0.05</td>
</tr>
<tr>
<td>407</td>
<td>2</td>
<td>1761</td>
<td>60</td>
<td>0.10</td>
</tr>
</tbody>
</table>

ORDERS_REL

<table>
<thead>
<tr>
<th>ORDER_NO</th>
<th>CUST_NO</th>
<th>ORDER_DATE</th>
<th>SHIP_DATE</th>
<th>TO_STREET</th>
<th>TO_CITY</th>
<th>TO_STATE</th>
<th>TO_ZIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>405</td>
<td>193</td>
<td>01/07/05</td>
<td>05/07/05</td>
<td>8 Automation Ln</td>
<td>Albany</td>
<td>NY</td>
<td>12205</td>
</tr>
<tr>
<td>407</td>
<td>195</td>
<td>29/06/05</td>
<td>01/07/05</td>
<td>400 E Joppa Rd</td>
<td>Baltimore</td>
<td>MD</td>
<td>21286</td>
</tr>
<tr>
<td>408</td>
<td>195</td>
<td>24/05/04</td>
<td>25/05/04</td>
<td>23985 Bedford Rd N</td>
<td>Battle Creek</td>
<td>MI</td>
<td>49017</td>
</tr>
</tbody>
</table>

Figure 4 – An instance of ORDERS_DB

```
<PurchaseOrder ID="405"/>
<OrderDate>2005-07-01</OrderDate>
<Customer>
<Name>Bryan Huston</Name>
<Address>
<Street>8 Automation Ln</Street>
<City>Albany</City>
<State>NY</State>
<ZIP>12205</ZIP>
</Address>
<Phone>+91 11 012 4813</Phone>

<LineItem>
<ItemNo>1</ItemNo>
<Product>
{Name>HD 10GB 5400</Name>
<Price>125.50</Price>
<TaxRate>0.01</TaxRate>
</Product>
<Quantity>35</Quantity>
<Discount>0.07</Discount>
</LineItem>

<PurchaseOrder ID="407"/>
<OrderDate>2005-06-29</OrderDate>
<Customer>
<Name>Cary Stockwell</Name>
<Address>
<Street>400 E Joppa Rd</Street>
<City>Baltimore</City>
<State>MD</State>
<ZIP>21286</ZIP>
</Address>
<Phone>+91 11 012 4835</Phone>

<LineItem>
<ItemNo>1</ItemNo>
<Product>
{Name>PC Bag - L/S</Name>
<Price>256.28</Price>
<TaxRate>0.05</TaxRate>
</Product>
<Quantity>15</Quantity>
<Discount>0.05</Discount>
</LineItem>

<PurchaseOrder ID="408"/>
<OrderDate>2004-05-24</OrderDate>
<Customer>
<Name>Cary Stockwell</Name>
<Address>
<Street>400 E Joppa Rd</Street>
<City>Baltimore</City>
<State>MD</State>
<ZIP>21286</ZIP>
</Address>
<Phone>+91 11 012 4835</Phone>

<LineItem>
<ItemNo>1</ItemNo>
<Product>
{Name>Mouse +WP/CL</Name>
<Price>32.89</Price>
<TaxRate>0</TaxRate>
</Product>
<Quantity>60</Quantity>
<Discount>0.1</Discount>
</LineItem>

<PurchaseOrder/>
```

Figure 5 – An instance of PurchaseOrder_XML view
In more detail, consider the instance (or database state) of ORDERS_DB shown in Figure 4. The corresponding instance of PurchaseOrder_XML is shown in Figure 5. This view instance contains a sequence of <PurchaseOrder> elements of type PurchaseOrder_Type, which are the primary elements of the view. Each <PurchaseOrder> element is constructed from a tuple of the ORDERS_REL table by using the SQL/XML publishing function XMLElement(). Function XMLElement() takes as arguments an element name, an optional collection of attributes, and zero or more additional arguments that make up the element content.

The sub-elements and attributes of a <PurchaseOrder> element are constructed by using SQL/XML sub-queries. For example,

- attribute ID is constructed using the subquery in line 4. Function XMLAttributes() produces, from its arguments, the attributes of its owner XMLElement() function. These arguments are value expressions to be evaluated, with optional aliases. The datatype of an attribute value expression cannot be an object type or a collection. If an attribute value expression evaluates to NULL, then no corresponding attribute is created.

- sub-element <Date> is constructed using the subquery in line 5. Function XMLForest() produces a forest of XML elements from its arguments, which are expressions to be evaluated, with optional aliases. If an expression evaluates to NULL, then no corresponding element is created.

- sub-element <LineItem> is constructed by the subquery in lines 18 to 29. Function XMLAgg() is an aggregate function that produces a forest of XML elements from a collection of XML elements where NULL arguments are dropped from the result. In lines 18 to 29, we have that, for each tuple in ORDERS_REL table, the relevant tuples of the LINE_ITEMS_REL table are retrieved and converted into a sequence of <LineItem> elements.

3. Basic Definitions

In this section, let $R, R_1,...,R_n$ be relation schemes of a relational schema $S$. Let $R, R_1,...,R_n$ be relations over $R, R_1,...,R_n$, respectively.

**Definition 1:** Let $f \subseteq R_1$ be a foreign key of $R_1$ that references $R_2$. Then, we say that:

i) $f$ is a link from $R_1$ to $R_2$.

ii) $f^{-1}$, the inverse of a $f$, is a link from $R_2$ to $R_1$. □

**Definition 2:**

i) Let $\ell$ be a link from $R_1$ to $R_2$ of the form $R_1[\alpha_1,...,\alpha_m] \sqsubseteq R_2[\beta_1,...,\beta_n]$. Let $r_1$ be a tuple of $R_1$. Then, $r_1/\ell = \{ r_2 \in R_2 | r_1.\alpha_i = r_2.\beta_i, \text{ for } 1 \leq i \leq m \}$. Let $r_2$ be a tuple of $R_2$. Then, $r_2/\ell = \{ r_1 \in R_1 | r_1.\alpha_i = r_2.\beta_i, \text{ for } 1 \leq i \leq m \}$. □

**Definition 3:** Let $\ell$ be a link from $R_1$ to $R_2$, and $r_1$ and $r_2$ be tuples of $R_1$ and $R_2$, respectively. Then, we say that:

i) $r_1$ references $r_2$ through $\ell$ iff $r_2 \in r_1/\ell$.

ii) $\ell$ has single occurrence iff a tuple of $R_1$ can reference at most one tuple of $R_2$ through $\ell$; otherwise, $\ell$ has multiple occurrence. □
Definition 4: Let $\ell_1, \ldots, \ell_n$ be links. Assume that:

i) $\ell_i$ is a foreign key of $R$ of the form $R[a_1', \ldots, a_m'] \subseteq R[b_1', \ldots, b_m']$ or the inverse of a foreign key of $R$ of the form $R[b_1', \ldots, b_m'] \subseteq R[a_1', \ldots, a_m']$

ii) $\ell_i$ is a foreign key of $R_{i-1}$ of the form $R_{i-1}[a_1', \ldots, a_m'] \subseteq R[b_1', \ldots, b_m']$ or the inverse of a foreign key of $R_i$ of the form $R[b_1', \ldots, b_m'] \subseteq R_{i-1}[a_1', \ldots, a_m']$, for $2 \leq i \leq n$.

Then, we say that:

i) $\phi = \ell_1, \ldots, \ell_n$ is a referential path from $R$ to $R_n$.

ii) the tuples of $R$ referential tuples of $R_n$ through $\phi$.

iii) $\phi$ has single occurrence iff $\ell_i$ has single occurrence, for $1 \leq i \leq n-1$; otherwise, $\phi$ has multiple occurrence.

Definition 5: Let $\phi = \ell_1, \ldots, \ell_n$ be a referential path from $R$ to $R_n$. Let $r$ be a tuple of $R$.

Then,

$r / \phi = \{ r \in R_n \mid (\exists r_1 \in R_1) \ldots (\exists r_{n-1} \in R_{n-1}) (r.a_1' = r_1.b_1', \text{ for } 1 \leq k \leq m) \text{ and } (r_i.a_i' = r_i.b_i', \text{ for } 1 \leq k \leq m_i \text{ and } 2 \leq i \leq n) \}$.

Definition 6: A path of $R$ is an expression of one of following forms:

i) NULL

ii) $a$, where $a$ is an attribute of $R$.

iii) $\{a_1, \ldots, a_n\}$, where $a_1, \ldots, a_n$ are attributes of $R$.

iv) $\phi.a$, where $\phi$ is a referential path from $R$ to $R'$ and $a$ is an attribute of $R'$.

v) $\phi.\{a_1, \ldots, a_n\}$, where $\phi$ is a referential path from $R$ to $R'$ and $a_1, \ldots, a_n$ are attributes of $R'$.

Definition 7: Let $r$ be a tuple of $R$.

i) $r / \text{NULL} = \{ r \}$.

ii) $r / a = \{ v \mid v = r.a \text{ and } v \neq \text{NULL} \}$, where $a$ is an attribute of $R$.

iii) $r / \{a_1, \ldots, a_n\} = \{ v \mid v = r.a_i \text{ with } 1 \leq i \leq m \text{ and } v \neq \text{NULL} \}$, where $a_1, \ldots, a_n$ are attributes of $R$.

iv) $r / \phi.a = \{ v \mid \exists r' \in r / \phi \text{ and } v \in r'/a \}$, where $\phi$ is a referential path from $R$ to $R'$, $a$ is an attribute of $R'$ and $r'$ is a tuple of $R'$.

v) $r / \phi.\{a_1, \ldots, a_n\} = \{ v \mid \exists r' \in r / \phi \text{ and } v \in r'/\{a_1, \ldots, a_n\} \}$, where $\phi$ is a referential path from $R$ to $R'$, $a_1, \ldots, a_n$ are attributes of $R'$ and $r'$ is a tuple of $R'$.

We say that an XML Schema complex type $T$ is restricted iff $T$ is defined using the complexType and sequence constructors only, and the type of its attributes is an XML simple type.

In the rest of this section, let $T$ be a restricted XML Schema complex type, and let $R$ and $R'$ be relation schemes of a relational schema $S$.

Definition 8: A correspondence assertion (CA) is an expression of the form $[T/e] \equiv [R/\delta]$ where $e$ is an element or an attribute of $T$, with type $T_e$, and $\delta$ is a path of $R$ such that:

i) If $e$ is an attribute or a single occurrence element and $T_e$ is a simple type, then $\delta$ has one of the following forms:

- $a$, where $a$ is an attribute of $R$ whose type is compatible with $T_e$;
- $\phi.a$, where $\phi$ is a referential path from $R$ to $R'$ such that $\phi$ has single occurrence, and $a$ is an attribute of $R'$ whose type is compatible with $T_e$.
ii) If \( \theta \) is a multiple occurrence element and \( T_\theta \) is an simple type, then \( \delta \) has one of the following forms:
- \( \varphi.a \), where \( \varphi \) is a referential path from \( R \) to \( R' \) such that \( \varphi \) has multiple occurrence and \( a \) is an attribute of \( R' \), whose type is compatible with \( T_\theta \);
- \( \{a_1,...,a_n\} \), where \( a_1,...,a_n \) are attributes of \( R \) such that the type of \( a_i \) is compatible with \( T_\theta \), for \( 1 \leq i \leq n \);
- \( \varphi\{a_1,...,a_n\} \), where \( \varphi \) is a referential path from \( R \) to \( R' \) such that \( \varphi \) has single occurrence, and \( a_1,...,a_n \) are attributes of \( R' \) such that the type of \( a_i \) is compatible with \( T_\theta \), for \( 1 \leq i \leq n \).

iii) If \( \theta \) is a single occurrence element and \( T_\theta \) is a complex type, then \( \delta \) has one of the following forms:
- \( \varphi \), where \( \varphi \) is a referential path from \( R \) to \( R' \) such that \( \varphi \) has single occurrence;
- \( \text{NULL} \)

iv) If \( \theta \) is a multiple occurrence element and \( T_\theta \) is a complex type, then \( \delta \) is a path from \( R \) to \( R' \) such that \( \delta \) has multiple occurrence.

**Definition 9:** Let \( \mathcal{A} \) be a set of correspondence assertions. We say that \( \mathcal{A} \) fully specifies \( T \) in terms of \( R \) iff
i) For each element or attribute \( \theta \) of \( T \), there is a single CA of the form \([T/\theta] = [R/\delta]\) in \( \mathcal{A} \), called the CA for \( \theta \) in \( \mathcal{A} \).
ii) For each assertion in \( \mathcal{A} \) of the form \([T/\theta] = [R/\delta]\), where \( \theta \) is an element of complex type \( T_\theta \) and \( \delta \) is a referential path from \( R \) to \( R' \), then \( \mathcal{A} \) fully specifies \( T_\theta \) in terms of \( R' \).
iii) For each assertion in \( \mathcal{A} \) of the form \([T/\theta] = [R/\text{NULL}]\), where \( \theta \) is an element of complex type \( T_\theta \), then \( \mathcal{A} \) fully specifies \( T_\theta \) in terms of \( R \). \( \square \)

**Definition 10:** Let \( \mathcal{A} \) be a set of correspondence assertions such that \( \mathcal{A} \) fully specifies \( T \) in terms of \( R \). Let \( R \) be a relation over \( R \).

i) Let \( S_1 \) be a set of elements that are instances of an XML simple type \( T \). Let \( S_2 \) be a set of values of an SQL scalar data type. We say that \( S_1 \equiv_\mathcal{A} S_2 \) iff
\[ \forall t \in S_1 \text{ there is } v \in S_2 \text{ such that } \text{f}(t) = \text{f}(v) \]
where \( f \) is a function that maps an SQL value to an XML value [10].

ii) Let \( S_1 \) be a set of values of an XML simple type. Let \( S_2 \) be a set of values of an SQL scalar data type. We say that \( S_1 \equiv_\mathcal{A} S_2 \) iff
\[ \forall v_1 \in S_1 \text{ there is } v_2 \in S_2 \text{ such that } v_1 = \text{f}(v_2) \]
where \( f \) is a function that maps an XML value to an SQL value [10].

iii) Let \( S_1 \) be a set of elements of an XML Schema complex type \( T \). Let \( S_2 \) be a set of tuples of \( R \). We say that \( S_1 \equiv_\mathcal{A} S_2 \) iff
\[ \forall t \in S_1 \text{ there is } r \in S_2 \text{ such that } t \equiv_r r. \]

iv) Let \( r \) be a tuple of \( R \) and let \( s \) be an instance of \( T \). We say that \( s \equiv_\mathcal{A} r \) iff, for each element \( \theta \) of \( T \) such that \([T/\theta] = [R/\delta]\) is the CA for \( \theta \) in \( \mathcal{A} \) (which exists by assumption on \( \mathcal{A} \)), then \([T/\theta] \equiv_\mathcal{A} r/\delta \), and, for each attribute \( a \) of \( T \) such that \([T/a] = [R/\delta]\) is the CA for \( a \) in \( \mathcal{A} \) (which exists by assumption on \( \mathcal{A} \)), then \( \text{DATA}(s/t/@a) \equiv_\mathcal{A} r/\delta \)
If \( s \equiv_\mathcal{A} r \), we say that \( s \) is semantically equivalent to \( r \) as specified by \( \mathcal{A} \). \( \square \)
4. Specifying XML Views

We propose to specify an XML view with the help of a set of correspondence assertions [12], which axiomatically specify how the XML view elements are synthesized from tuples of the base source. Let $S$ be the base relational schema. An XML view, or simply, a view over $S$ is a quadruple $V=<e, T, R, A>$, where:

(i) $e$ is the name of the primary element of the view;
(ii) $T$ is the XML type of element $e$;
(iii) $R$ is a relation scheme or a relational view scheme of $S$;
(iv) $A$ is a set of path correspondence assertions that fully specifies $T$ in terms of $R$.

We say that the pair $<e, T>$ is the view schema of $V$ and that $R$ is the pivot relation scheme of the view.

Consider, for example, the view $\text{PurchaseOrder}_\text{XML}$, whose primary element $<\text{PurchaseOrder}>$ has type $\text{PurchaseOrder}_\text{Type}$, and whose pivot relation scheme is $\text{ORDERS}_\text{REL}$. Figure 3 shows an SQL/XML specification of $\text{PurchaseOrder}_\text{XML}$ and Figure 2 depicts a graphical representation of $\text{PurchaseOrder}_\text{Type}$. Figure 6 shows the correspondence assertions of $\text{PurchaseOrder}_\text{XML}$, which fully specifies $\text{PurchaseOrder}_\text{Type}$ in terms of $\text{ORDERS}_\text{REL}$.

We developed a tool, called XML View-By-Assertions (XVBA), to support the definition of view correspondence assertions. XVBA features a simple graphical interface which allows the user to navigate to related tables. The process starts with the user

```
<table>
<thead>
<tr>
<th>PurchaseOrder_Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>@ID</td>
</tr>
<tr>
<td>OrderDate</td>
</tr>
<tr>
<td>Customer</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Phone*</td>
</tr>
<tr>
<td>ItemNo</td>
</tr>
<tr>
<td>Product</td>
</tr>
<tr>
<td>Price</td>
</tr>
<tr>
<td>TaxRate</td>
</tr>
<tr>
<td>Quantity</td>
</tr>
<tr>
<td>Discount</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ORDERS_REL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORDER_NO</td>
</tr>
<tr>
<td>CUST_NO</td>
</tr>
<tr>
<td>ORDER_DATE</td>
</tr>
<tr>
<td>FK1 (CUSTOMERS_REL)</td>
</tr>
<tr>
<td>CUST_NO</td>
</tr>
<tr>
<td>CUST_NAME</td>
</tr>
<tr>
<td>STREET</td>
</tr>
<tr>
<td>CITY</td>
</tr>
<tr>
<td>STATE</td>
</tr>
<tr>
<td>ZIP</td>
</tr>
<tr>
<td>PHONE1</td>
</tr>
<tr>
<td>PHONE2</td>
</tr>
<tr>
<td>PHONE3</td>
</tr>
</tbody>
</table>

Figure 6 – Correspondence Assertions of PurchaseOrder_XML view
```
loading a source and view schemas into XVBA. The user can then graphically connect elements or attributes of the XML type with attributes or paths of the pivot relation.

The correspondence assertions of PurchaseOrder_XML are generated by: (1) matching the elements and attributes of PurchaseOrder_Type with attributes or paths of ORDERS_REL; and (2) recursively descending into sub-elements of PurchaseOrder_Type to define their correspondence assertion. For example, to define the assertion of the element LineItem ($L_1$,[PurchaseOrder_Type/LineItems] $\equiv$ [ORDERS_REL/ORDERS_REL$^{-1}$]), the user selects the element LineItem on the view schema and the inverse foreign key FK2$^{-1}$ on the database schema.

5. Mapping Assertions to SQL/XML

Let $S$ be a relational schema and $V=< e, T, R, A >$ be a XML view over $S$. In this section, we show that the view correspondence assertions in $A$ define a mapping that can be correctly translated to an SQL/XML query view definition.

Given an instance $\sigma_S$ of $S$, let $\sigma_S(R)$ denote the relation that $\sigma_S$ associates with $R$. Moreover, given an element $e$, the extended content of $e$ is the list of attributes and child elements of $e$. The correspondence assertions in $A$ define a functional mapping, denoted $\text{DEF}_V$, from instances of the source schema $S$ to instances of the view schema. Given an instance $\sigma_S$ of $S$, the value of $V$ on $\sigma_S$ is given by:

$$\text{DEF}_V(\sigma_S) = \{ s | s \text{ is an } <e> \text{ element of type } T \text{ and } \exists r \in \sigma_S(R) \text{ such that } s \equiv_r r \}$$

The SQL/XML definition of $V$ is given by:

```sql
CREATE VIEW V OF XMLTYPE
AS SELECT XMLELEMENT( "e", \tau[R\rightarrow T](r) )
FROM R r
```

where $\tau[R\rightarrow T](r)$ is a sequence of SQL/XML sub-queries, one for each element or attribute of $T$. Given a tuple $r$ of $\sigma_S(R)$, $\sigma_S(\tau[R\rightarrow T](r))(r)$ denotes the result of evaluating the SQL/XML sub-queries in the instance $\sigma_S$, with $r$ replaced by $r$. We will prove that, given an instance $St$ of $T$ whose extended content is constructed from $\sigma_S(\tau[R\rightarrow T](r))(r)$, then $St \equiv_r r$.

Figure 7 presents the algorithm $\text{GenConstructor}$ that generates the constructor function $\tau[R\rightarrow T](r)$. Figure 8 presents the algorithm $\text{GenSQL/XMLSubquery}$, where $\phi$ is a path of the form $\ell_1$, ... $\ell_n$, as defined in Definition 6, and $\text{Join}\phi(r)$ is defined by the following SQL fragment:

$$R_1 r_1, ..., R_n r_n$$

```sql
WHERE r_1.a_{l_1} = r_1.b_{l_1} AND ... AND r_n.a_{l_n} = r_n.b_{l_n}
AND r_1.a_{l_2} = r_2.b_{l_2} AND ... AND r_n.a_{l_n} = r_n.b_{l_n}
```

Figure 7 presents the algorithm $\text{GenConstructor}$ that generates the constructor function $\tau[R\rightarrow T](r)$. Figure 8 presents the algorithm $\text{GenSQL/XMLSubquery}$, where $\phi$ is a path of the form $\ell_1$, ... $\ell_n$, as defined in Definition 6, and $\text{Join}\phi(r)$ is defined by the following SQL fragment:

```sql
WHERE r_1.a_{l_1} = r_1.b_{l_1} AND ... AND r_n.a_{l_n} = r_n.b_{l_n}
AND r_1.a_{l_2} = r_2.b_{l_2} AND ... AND r_n.a_{l_n} = r_n.b_{l_n}
```
Input: a XML Type \( T \), a relation scheme \( R \), a set of correspondence assertions \( \mathcal{A} \) that fully specifies \( T \) in terms of \( R \) and an alias \( r \) for \( R \).

Output: Function \( \tau[R \rightarrow T][r] \).

Let \( \tau \) be an string;

\( \tau := \emptyset \);

If \( T \) has attributes then

\( \tau := \tau + "XMLAttributes(" \);

For each attribute \( a \) of type \( T \) where \( \Psi_a \) is the CA for \( a \) in \( \mathcal{A} \) do

\( \tau := \tau + \text{GenSQL/XMLSubquery}(\mathcal{A}, \Psi_a, r) \);

end for;

\( \tau := \tau + ")" \)

End If;

For each element \( e \) of \( T \) where \( \Psi_e \) is the CA for \( e \) in \( \mathcal{A} \) do

\( \tau := \tau + \text{GenSQL/XMLSubquery}(\mathcal{A}, \Psi_e, r) \);

End for;

Return \( \tau \);

---

**Figure 7 – Algorithm GenConstructor**

Input: a set of correspondence assertions \( \mathcal{A} \) that fully specifies \( T \) in terms of \( R \), the CA \( [T/e] = [R/\delta] \) in \( \mathcal{A} \) where \( e \) is an element or attribute of type \( T_e \), and an alias \( r \) for \( R \).

Output: a SQL/XML sub-query

Let \( Q \) be an string;

In case of

Case 1: If \( e \) is a single occurrence element, \( T_e \) is a simple type and \( \delta = a \), then

\( Q := "XMLFOREST(r.a AS \"e\")"; \)

Case 2: If \( e \) is a single occurrence element, \( T_e \) is a simple type and \( \delta = \varphi.a \), then

\( Q := "XMLFOREST((SELECT r_n.a FROM Join(\varphi(r))) AS \"e\")"; \)

Case 3: If \( e \) is a multiple occurrence element, \( T_e \) is a simple type and \( \delta = \{a_1, \ldots, a_n\} \), then

\( Q := "XMLCONCAT( XMLFOREST(r.a AS \"e\"), \ldots, XMLFOREST(r.a AS \"e\"))"; \)

Case 4: If \( e \) is a multiple occurrence element, \( T_e \) is a simple type and \( \delta = \varphi \{a_1, \ldots, a_n\} \), then

\( Q := "XMLCONCAT( (SELECT XMLFOREST(r_n.a AS \"e\") FROM Join(\varphi(r))) )"; \)

Case 5: If \( e \) is a multiple occurrence element, \( T_e \) is a simple type and \( \delta = \varphi.a \), then

\( Q := "(SELECT XMLAGG( XMLFOREST(r_n.a AS \"e\") FROM Join(\varphi(r))) )"; \)

Case 6: If \( e \) is a single occurrence element, \( T_e \) is a complex type and \( \delta = \varphi \), then

\( Q := "(SELECT XMLELEMENT("e", GenConstructor(T_e, R_n, A, r_n) + )) FROM Join(\varphi(r)))"; \)

Case 7: If \( e \) is a multiple occurrence element, \( T_e \) is a complex type and \( \delta = \varphi \), then

\( Q := "(SELECT XMLELEMENT("e", " + GenConstructor(T_e, R_n, A, r_n) + ") FROM Join(\varphi(r)))"; \)

Case 8: If \( e \) is a single occurrence element, \( T_e \) is a complex type and \( \delta = NULL \), then

\( Q := "XMLELEMENT("e", " + GenConstructor(T_e, R_n, A, r_n) + ")"; \)

Case 9: If \( e \) is an attribute, \( T_e \) is a simple type and \( \delta = a \), then

\( Q := "r.a AS \"e\"; \)

Case 10: If \( e \) is an attribute, \( T_e \) is an simple type and \( \delta = \varphi.a \), then

\( Q := "(SELECT r_n.a FROM Join(\varphi(r))) AS \"e\"; \)

End case;

return \( Q \);

---

**Figure 8 – Algorithm GenSQL/XMLSubquery**
The correctness of these algorithms follows from the propositions and theorem below. In what follows, let \( T \) be a XML Schema type, \( R \) be a relation scheme, \( \mathcal{A} \) be a set of correspondence assertions that fully specifies \( T \) in terms of \( R \), and \( t \) be an alias for \( R \).

**Proposition 1:** Let \( \Psi \) be the CA \( [T/\theta] \equiv [R/\delta] \) for element \( \theta \) of type \( T_\theta \) in \( \mathcal{A} \). Let \( \text{GenSQL/XMLSubquery}(\mathcal{A}, \Psi, r) = Q_\theta[r] \). Let \( r \) be a tuple of \( \sigma_S(R) \) and \( \mathcal{S} \) be the set of \( \angle \theta \) elements resulting from evaluating \( Q_\theta[r] \) in \( \sigma_S \) with \( r \) replaced by \( r \). Then, we have that \( \mathcal{S} \equiv_T r/\delta \). \( \square \) (See [13] for the proof).

**Proposition 2:** Let \( \Psi \) be the CA \( [T/\alpha] \equiv [R/\delta] \) for attribute \( \alpha \) of type \( T_\alpha \) in \( \mathcal{A} \). Let \( \text{GenSQL/XMLSubquery}(\mathcal{A}, \Psi, r) = Q_\alpha[r] \). Let \( r \) be a tuple of \( \sigma_S(R) \) and \( \alpha \) be the value resulting from evaluating \( Q_\alpha[r] \) in \( \sigma_S \) with \( r \) replaced by \( r \). Then, we have that \( \alpha = f(\psi) \), where \( \psi \) is the only value in \( r/\delta \), and \( f \) is a function that maps SQL values to XML values [5]. \( \square \) (See [13] for the proof).

**Theorem 1:**

Let \( a_1, \ldots, a_k \) be the attributes of \( T \) and let \( e_1, \ldots, e_m \) be the elements of \( T \).

Let \( \text{GenConstructor}(R, t, T, \mathcal{A}) = \tau[R\rightarrow T][r] \)

Let \( t \) be a tuple of \( \sigma_S(R) \).

Let \( S \) be an \( \angle \theta \) element of type \( T \) whose extended content is constructed from \( \sigma_S(\tau[R\rightarrow T][r])(r) \).

Then, \( S \equiv_T t \).

**Proof:** Let \( a_1, \ldots, a_k \) be the attributes of \( T \). Let \( \Psi_{\alpha_i} \) be the CA for \( a_i \) in \( \mathcal{A} \) and \( T_\alpha \) be the type of \( a_i \) for \( 1 \leq i \leq k \). Assume that \( \Psi_{\alpha_i} \) is of the form \( [T/\alpha] \equiv [R/\delta_{\alpha_i}] \). Let \( e_1, \ldots, e_m \) be the elements of \( T \). Let \( \Psi_{e_i} \) be the CA for \( e_i \) in \( \mathcal{A} \) and \( T_{e_i} \) be the type of \( e_i \) for \( 1 \leq i \leq m \). Assume that \( \Psi_{e_i} \) is of the form \( [T/e] \equiv [R/\delta_{e_i}] \).

Let \( \tau[R\rightarrow T][r] \) be the constructor function generated by \( \text{GenConstructor} \). From the algorithm, we have that:

\[
\tau[R\rightarrow T][r] = \text{XMLAttributes}(Q_{\alpha_1}[r], \ldots, Q_{\alpha_k}[r], Q_{e_1}[r], \ldots, Q_{e_m}[r]),
\]

where

\[
Q_{\alpha_i}[r] = \text{GenSQL/XMLSubQuery}(\mathcal{A}, \Psi_{\alpha_i}, r), \quad \text{for } 1 \leq i \leq k
\]

\[
Q_{e_i}[r] = \text{GenSQL/XMLSubQuery}(\mathcal{A}, \Psi_{e_i}, r), \quad \text{for } 1 \leq i \leq m
\]

Let \( r \) be a tuple of \( \sigma_S(R) \). Let \( S \) be an \( \angle \theta \) element of type \( T \) whose extended content is constructed from \( \sigma_S(\tau[R\rightarrow T][r])(r) \). For \( 1 \leq i \leq k \), let \( \alpha_i \) be the value resulting from evaluating \( Q_{\alpha_i}[r] \) in \( \sigma_S \) with \( r \) replaced by \( r \). For \( 1 \leq i \leq m \), let \( S_{e_i} \) be the set of \( \angle \theta \) elements resulting from evaluating \( Q_{e_i}[r] \) in \( \sigma_S \) with \( r \) replaced by \( r \). Therefore, \( S = \langle e_1 = "\alpha_1" \ldots a_k = "\alpha_k" \rangle S_{e_1} \ldots S_{e_m} \rangle \langle /e \rangle \). From Proposition 1, we have \( S_{e_i} \equiv_T r/\delta_{e_i} \), for \( 1 \leq i \leq m \). From Proposition 2, we have \( \alpha_i = f(\psi) \), for \( 1 \leq i \leq k \), where \( \psi \) is the only value in \( r/\delta_{\alpha_i} \), and \( f \) is a function that maps SQL values to XML values. So, from Definition 10 (ii), we have that \( \text{DATA}(S) = f(\psi) \), for \( 1 \leq i \leq k \). Therefore, from Definition 10 (iv), we have that \( S \equiv_T t \).

In what follows, for simplicity, let \( \tau[R\rightarrow T][r] \) denote the function \( \sigma_S(\tau[R\rightarrow T][r])(r) \) that constructs the extended content of an instance \( S \) of \( T \) such that \( S \equiv_T t \).

Consider, for example, the SQL/XML definition of \textbf{PurchaseOrder XML}, shown in Figure 3. The constructor function \( \tau[\text{ORDERS}_{\text{REL}}\rightarrow \text{PurchaseOrder}_{\text{Type}}](O) \) (lines 3 to 29) constructs the extended content of an instance of \textbf{PurchaseOrder Type} from a tuple.
of ORDERS_REL. The constructor function contains four sub-queries, one for each
element or attribute of PurchaseOrder_Type. In GenSQL/XMLSubquery, each subquery is
generated from the CA of the corresponding element or attribute. Figure 3 shows the
assertion that generates each SQL/XML subquery of $\tau[ORDERS_REL\rightarrow PurchaseOrder_Type](O)$.

We will show that $\tau[ORDERS_REL\rightarrow PurchaseOrder_Type](O)$ constructs the
extended content of an instance $SP$ of PurchaseOrder_Type, for each tuple $O$ of an
instance of ORDERS_REL, such that $SP$ is semantically equivalent to $O$, as specified by
the assertions of PurchaseOrder_XML.

Let ORDERS_DB be an instance of the relational schema ORDERS_DB. Let
CUSTOMER_REL, ORDERS_REL, PRODUCTS_REL and LINE_ITEMS_REL be the
instances that ORDERS_DB associates with the relation schemes CUSTOMER_REL,
ORDERS_REL, PRODUCTS_REL and LINE_ITEMS_REL of ORDERS_DB, respectively.

Let $O$ be a tuple of ORDERS_REL and let $SP$ be an instance of PurchaseOrder_Type whose extended content is constructed with $\tau[ORDERS_REL\rightarrow PurchaseOrder_Type](O)$. From Definition 10 and from $\Psi_1, \Psi_2, \Psi_3$ and $\Psi_{11}$, we have that $SP \equiv_{A} O$ iff:

1. DATA($SP / @$ID) $\equiv_{A} O / ORDER_NO$,
2. $SP / OrderDate $\equiv_{A} O / ORDER_DATE$,
3. $SP / Customer $\equiv_{A} O / FK1$, and
4. $SP / Lineltem $\equiv_{A} O / FK2^{-1}$.

Proof of (1): From line 4 of Figure 3, we have that:
DATA($SP / @$ID) = \{ v \mid v = f(O. ORDER_NO) and v \neq NULL \}.
From Definition 7 (ii), we have that $O / ORDER_NO = \{ D \mid D = O. ORDER_NO and O. ORDER_NO \neq NULL \}.
So, from Definition 10 (i), we have that: DATA($SP / @$ID) $\equiv_{A} O / ORDER_NO$.

Proof of (2): From line 5 of Figure 3, we have that:
$SP / OrderDate = \{ SD \mid SD = \langle ORDER_DATE \rangle f(O. ORDER_DATE) \langle ORDER_DATE \rangle and O. ORDER_DATE \neq NULL \}.
From Definition 7 (ii), we have that:
$O / ORDER_DATE = \{ D \mid D = O. ORDER_DATE and O. ORDER_DATE \neq NULL \}.
So, from Definition 10 (i), we have that: $SP / OrderDate $\equiv_{A} O / ORDER_DATE$.

Proof of (3): From lines 6-17 of Figure 3, we have that:
$SP / Customer = \{ SC \mid \exists C \in CUSTOMER_REL such that
C.CUST_NO = O.CUST_NO and the extended content of SC is
constructed from $\tau[CUSTOMERS_REL\rightarrow Customer_Type](C)$.
In following, we show that, given a tuple $C \in CUSTOMER_REL$ and an element $SC$ whose extended content is constructed from $\tau[CUSTOMERS_REL\rightarrow Customer_Type](C)$, then $SC \equiv_{A} C$. From $\Psi_4, \Psi_5, \epsilon \Psi_{10}$, we have that $SC \equiv_{A} C$ iff:
(3.1) $SC / Name \equiv_{A} C / CUST_NAME$,
(3.2) $SC / Address \equiv_{A} C / NULL$, and
(3.3) $SC / Phone \equiv_{A} C / \{PHONE1, PHONE2, PHONE3\}$. 

Proof of (3.1): The proof follows from line 7 of Figure 3 and is similar to the proof of (2).

Proof of (3.2): From lines 8-12 of Figure 3, we have that:
$\text{SC} / \text{Address} = \{ \text{SA} \}$ where the extended content of $\text{SA}$ is constructed from $\tau[\text{CUSTOMERS_REL} \rightarrow \text{Address_Type}] (C)$.

In following, we show that, if the extended content of $\text{SA}$ is constructed from $\tau[\text{CUSTOMERS_REL} \rightarrow \text{Address_Type}] (C)$ then $\text{SA} \equiv \text{C}$. From $\Psi_6, \Psi_7, \Psi_8$ and $\Psi_9$, we have that $\text{SA} \equiv \text{C}$ iff

1. $\text{SA} / \text{Street} \equiv \text{C} / \text{STREET}$,
2. $\text{SA} / \text{City} \equiv \text{C} / \text{CITY}$,
3. $\text{SA} / \text{State} \equiv \text{C} / \text{STATE}$, and
4. $\text{SA} / \text{ZIP} \equiv \text{C} / \text{ZIP}$.

The proofs of (3.2.1), (3.2.2), (3.2.3) and (3.2.4) are similar to the proof of (2) and follow from lines 9, 10, 11 and 12 of Figure 3, respectively. So, from (3.2.1), (3.2.2), (3.2.3) and (3.2.4), we have that $\text{SC} / \text{Address} = \{ \text{SA} \}$ where $\text{SA} \equiv \text{C}$.

From Definition 7 (i), we have that $\text{C} / \text{NULL} = \{ \text{C} \}$. Therefore, from Definition 10 (iii), we have that $\text{SC} / \text{Address} \equiv \text{C} / \text{NULL}$.

Proof of (3.3): From lines 13-15 of Figure 3, we have that
$\text{SC} / \text{Phone} = \text{S1} \cup \text{S2} \cup \text{S3}$ where:

- $\text{S1} = \{ \text{H} | \text{H} = \langle \text{Phone} \rangle \langle \text{C}.\text{PHONE1} \rangle \langle \text{Phone} \rangle \text{ and } \text{H} \neq \text{NULL} \}$,
- $\text{S2} = \{ \text{H} | \text{H} = \langle \text{Phone} \rangle \langle \text{C}.\text{PHONE2} \rangle \langle \text{Phone} \rangle \text{ and } \text{H} \neq \text{NULL} \}$, and
- $\text{S3} = \{ \text{H} | \text{H} = \langle \text{Phone} \rangle \langle \text{C}.\text{PHONE3} \rangle \langle \text{Phone} \rangle \text{ and } \text{H} \neq \text{NULL} \}$.

From Definition 7 (iii) we have that:
$\text{C} / \{ \text{PHONE1, PHONE2, PHONE3} \} = \{ \text{H} | ( \text{H} = \text{C}.\text{PHONE1} \text{ or } \text{H} = \text{C}.\text{PHONE2} \text{ or } \text{H} = \text{C}.\text{PHONE3} ) \text{ and } \text{H} \neq \text{NULL} \}$.

Therefore, from Definition 10 (ii), we have that:
$\text{SC} / \text{Phone} \equiv \text{C} / \{ \text{PHONE1, PHONE2, PHONE3} \}$.

So, from (3.1), (3.2) and (3.3), we have that:
$\text{SP} / \text{Customer} = \{ \text{SC} | \exists \text{C} \in \text{CUSTOMER_REL} \text{ such that } \text{C.CUST_NO} = \text{O.CUST_NO} \text{ and } \text{SC} \equiv \text{C} \}$.

From Definition 2 (i), we have that:
$\text{O} / \text{FK1} = \{ \text{C} | \exists \text{C} \in \text{CUSTOMER_REL} \text{ such that } \text{C.CUST_NO} = \text{O.CUST_NO} \}$. Therefore, from Definition 10 (iii), we have that: $\text{SP} / \text{Customer} \equiv \text{O} / \text{FK1}$.

Proof of (4): From lines 18-29 of Figure 3, we have that:
$\text{SP} / \text{LineItem} = \{ \text{SL} | \exists \text{L} \in \text{ITEMS_REL} \text{ such that } \text{L.ORDER_NO} = \text{O.ORDER_NO} \text{ and } \text{the extended content of } \text{SL} \text{ is constructed from } \tau[\text{LINE_ITEMS_REL} \rightarrow \text{LineItem_Type}] (\text{L}) \}.$

In following, we show that, given a tuple $\text{L} \in \text{LINE_ITEMS_REL}$ and an element $\text{SL}$ whose extended content is constructed from $\tau[\text{LINE_ITEMS_REL} \rightarrow \text{LineItem_Type}] (\text{L})$, then $\text{SL} \equiv \text{L}$. From $\Psi_{12}, \Psi_{13}, \Psi_{17}$ and $\Psi_{18}$, we have that $\text{SL} \equiv \text{L}$ iff:

1. $\text{SL} / \text{ItemNo} \equiv \text{L} / \text{ITEM_NO}$,
2. $\text{SL} / \text{Product} \equiv \text{L} / \text{FK3}$,
3. $\text{SL} / \text{Quantity} \equiv \text{L} / \text{QUANTITY}$, and
4. $\text{SL} / \text{Discount} \equiv \text{L} / \text{DISCOUNT}$.
The proofs of (4.1), (4.3) and (4.4) are similar to the proof of (2) and follow from lines 19, 26 and 27 of Figure 3, respectively.

**Proof of (4.2):** From lines 19-25, we have that
\[ SL / Product = \{ SD | \exists D \in PRODUCTS_REL \text{ such that } D.PROD_NO = L.PROD_NO \text{ and } \]
\[ \text{the extended content of } SD \text{ is constructed from } \tau[PRODUCTS_REL \rightarrow Product_Type](D) \}. \]

In following, we show that, given a tuple \( D \in PRODUCTS_REL \) and an element \( SD \) whose extended content is constructed from \( \tau[PRODUCTS_REL \rightarrow Product_Type](D) \), then \( SD \equiv_{\lambda} D \). From \( \Psi_{14}, \Psi_{15} \) and \( \Psi_{16} \), we have that \( SD \equiv_{\lambda} D \) iff:

- (4.2.1) \( SD / Name \equiv_{\lambda} D / NAME \),
- (4.2.2) \( SD / Price \equiv_{\lambda} D / PRICE \), and
- (4.2.3) \( SD / TaxRate \equiv_{\lambda} D / TAX_RATE \).

The proofs of (4.2.1), (4.2.2) and (4.2.3) are similar to the proof of (2) and follow from lines 21, 22 and 23 of Figure 3, respectively. So, from (4.2.1), (4.2.2) and (4.2.3), we have that:

\[ SL / Product = \{ SD | \exists D \in PRODUCTS_REL \text{ such that } D.PROD_NO = L.PROD_NO \text{ and } SD \equiv_{\lambda} D \}. \]

From Definition 2 (i), we have that:

\[ L / FK3 = \{ D | \exists D \in PRODUCTS_REL \text{ such that } D.PROD_NO = L.PROD_NO \}. \]

Therefore, from Definition 10 (iii), we have that: \( SL / Product \equiv_{\lambda} L / FK3 \). So, from (4.1), (4.2), (4.3) and (4.4), we have that:

\[ SP / LineItem = \{ SL | \exists L \in ITEMS_REL \text{ such that } L.ORDER_NO = O.ORDER_NO \text{ and } \]
\[ SL \equiv_{\lambda} L \}. \]

From Definition 2 (ii), we have that:

\[ O / FK2^{-1} = \{ L | \exists L \in ITEMS_REL \text{ such that } L.ORDER_NO = O.ORDER_NO \}. \]

Therefore, from Definition 10 (iii), we have that: \( SP / LineItem \equiv_{\lambda} O / FK2^{-1} \).

6. Conclusions

We argued in this paper that we may fully specify an XML view in terms of a relational schema using view correspondence assertions, in the sense that the assertions define a mapping from instances of the relational schema to instances of the XML view schema.

We presented an algorithm that generates, based on the view correspondence assertions, the SQL/XML view definition. Moreover, we showed that the SQL/XML query generated by the algorithm correctly represents the mapping defined by the view correspondence assertions.

As future work, we envision a number of extensions to our mapping formalism to express broader types of view schema. We are working in generalizing our mapping formalism and algorithm to deal with XML views of object-relational data.

References


