Fuzzy Integrity Constraints for Native XML Database

E.J.Thomson Fredrick [1], Dr.G.Radhamani [2]

[1] Research Scholar, Research & Development Centre, Bharathiar University
Coimbatore – 641 046, Tamil Nadu, India

[2] Director & Professor, Dr.G.R.D College of Science
Coimbatore - 641 014, Tamil Nadu, India

Abstract
XML is gradually accepted as a standard for representing, accessing, and exchanging data in internet applications. It poses many new challenges to XML storages or XML repositories. The data centric approach of XML has necessitated the definition of integrity constraints for XML. Integrity constraints have proved fundamentally important in Relational Database management. We introduce Fuzzy Logic based constraints using triangular membership function and Trapezium shape membership functions. In this paper, we have introduced a fuzzy constraint-based framework to model the schema matching problem. Our approach is able to handle uncertainty in schema matching and data inconsistency in Native XML Database systems by exploiting fuzzy constraints. Our approach will restrict invalid XML data into XML Database by using fuzzy domain integrity constraints.

Keywords: *XML Schema, Fuzzy Sets, Fuzzy Constraints and Native XML Database System*

1 Introduction

XML is now readily used for the management and interchange of enterprise data. It is critical to many of the applications relying on XML data that integrity of data be maintained. XML Database vendors rushed to enrich their products with more flexible and advanced features to make them satisfy the requirements of modern applications. XML (*eXtensible Markup Language*) is a current standard for representation and interchange of data on the Web [XML 2005]. An XML document is a collection of data, like a database. However, XML technology is not equivalent to database technology, because there are not effective solutions for all aspects of XML data management, like integrity constraint control and transaction management [Chaudhry and Zicary 2003].

Integrity constraints for XML documents are mainly defined through XML schemas: DTD and XML Schema. However, XML schemas do not have support for all existing kinds of domain integrity constraints in SQL databases. Integrity constraints have proved fundamentally important in database management. (Yunsheng Liu *et al*, 2004) So in this paper, we propose integrity constraints based on the Fuzzy Logic. We propose a general approach for representing fuzzy information in XML documents by using Fuzzy Integrity constraints based XML Schema. Researchers have relaxed the classical relational model of storing data to incorporate imprecision and fuzziness. Two levels of work have been accomplished in the relational database area to include fuzziness. The first level is making imprecise or flexible queries on the classical precise relational data. The second level is related to adding imprecision in the stored data. (Abhishek Gaurav and Reda Alhajj,2006). In this paper, we describe the work done at the second level as we are concerned with incorporating fuzziness into XML Schema and XML documents. Our approach will enforce business rules in Native XML Database transactions.

2 Related Work

Buckles B. P. and Petry F. E, 1982 are among the pioneers in using fuzzy theory in the database area; they used the similarity relations to relax the relational model to model uncertain data. Klaus Turowski and Uwe Weng, 2002 introduced a formal syntax for important fuzzy data types used to store fuzzy information. They defined appropriate document type definitions (DTD), and showed how fuzzy information, whose description is based on these DTDs, can be exchanged between application systems by means of the extensible markup language (XML). Carlo Combi et al, 2004 proposed a flexible approach to evaluate fuzzy association rules on XML documents. In particular they described evaluation techniques in order to assign a degree to structural association rules, which allow one to evaluate the similarity of the XML document with respect to a given structure, and to value association rules which allow one to capture the similarity between the information contained in the XML document and the required information.

Abhishek Gaurav and Reda Alhajj,2006 defined the mechanism to represent fuzzy data along with crisp data in XML format. They proposed few additional tags in the XML schema to store the fuzzy data and the measurement of fuzziness. Barbara Oliboni and Gabriele Pozzani, 2008 tried to represent different aspects of fuzzy information by adapting a data type classification already proposed for the relational database context, and by integrating different kinds of fuzzy information to compose a complete definition. They proposed a general approach for representing fuzzy information in XML documents by using XML schema, and describe a fuzzy XML schema definition taking into account fuzzy data types and elements needed to fully represent fuzzy information. Alsayed Algergawy *et al*, 2008 built a conceptual connection between the *Schema Matching Problem* and the *fuzzy constraint optimization problem.* In particular, they proposed the use of the *fuzzy constraint optimization problem* as a framework to model and formalize the schema matching problem.

But in this paper, we concentrate only on how to introduce Domain Integrity Constraints using Fuzzy Sets for implementing data consistency in the Native XML Database Systems. Our approach will work similar to the Check constraint in Relational Database Management Systems. We implement Fuzzy triangular membership functions based rules for implementing Domain Integrity Constraints in Native XML Database Systems.

3 Technical Background

3.1 Brief Overview of Fuzzy Theory

The fuzzy set theory [1] is a generalization of the classical set theory in which a characteristic membership function is attached to a set (Zadeh L.A,1965). The membership of an object in a fuzzy set is a fuzzy concept, and an object belongs to a certain degree as opposed to classical set theory where an object either belongs or does not belong to a set.

Formally, consider fuzzy set *A*, its domain *D*, and object *x*. Membership function μ specifies the degree of membership of *x* in *A*, such that $\mu_A(x)$: D→ [0, 1]. $\mu_A(x) = 0$ means *x* does not belong to *A*, whereas $\mu_A(x) = 1$ means *x* completely belongs to *A*. Intermediate values represent varying degree of membership. For example, Fuzzy sets can represent the linguistic terms such as very tall, tall, medium tall, short and very short of Height criteria. A reasonable expression of these concepts by triangular membership functions H_1, H_2, H_3, H_4, H_5 is shown in Figure.1. These linguistic terms are defined on the interval [135,185] are as follows:

$$H_1 = \left\{ \begin{array}{ll} 0 & when\ \mathrm{x} <= 135 \\ (145 - x)/10 & when\ 135\ >\ \mathrm{x}\ < 145 \\ 1 & when\ \mathrm{x} >= 145 \end{array} \right\}$$

$$H_2 = \left\{ \begin{array}{ll} 0 & when\ \mathrm{x} <= 145 \\ (155 - x)/10 & when\ 145\ >\ \mathrm{x}\ < 155 \\ 1 & when\ \mathrm{x} >= 155 \end{array} \right\}$$

$$H_3 = \left\{ \begin{array}{ll} 0 & when\ \mathrm{x} <= 155 \\ (165 - x)/10 & when\ 155\ >\ \mathrm{x}\ < 165 \\ 1 & when\ \mathrm{x} >= 165 \end{array} \right\}$$

$$H_4 = \left\{ \begin{array}{ll} 0 & when\ \mathrm{x} <= 165 \\ (175 - x)/10 & when\ 165\ >\ \mathrm{x}\ < 175 \\ 1 & when\ \mathrm{x} >= 175 \end{array} \right\}$$

$$H_5 = \left\{ \begin{array}{ll} 0 & when\ \mathrm{x} <= 175 \\ (185 - x)/10 & when\ 175\ >\ \mathrm{x}\ < 185 \\ 1 & when\ \mathrm{x} >= 185 \end{array} \right\}$$

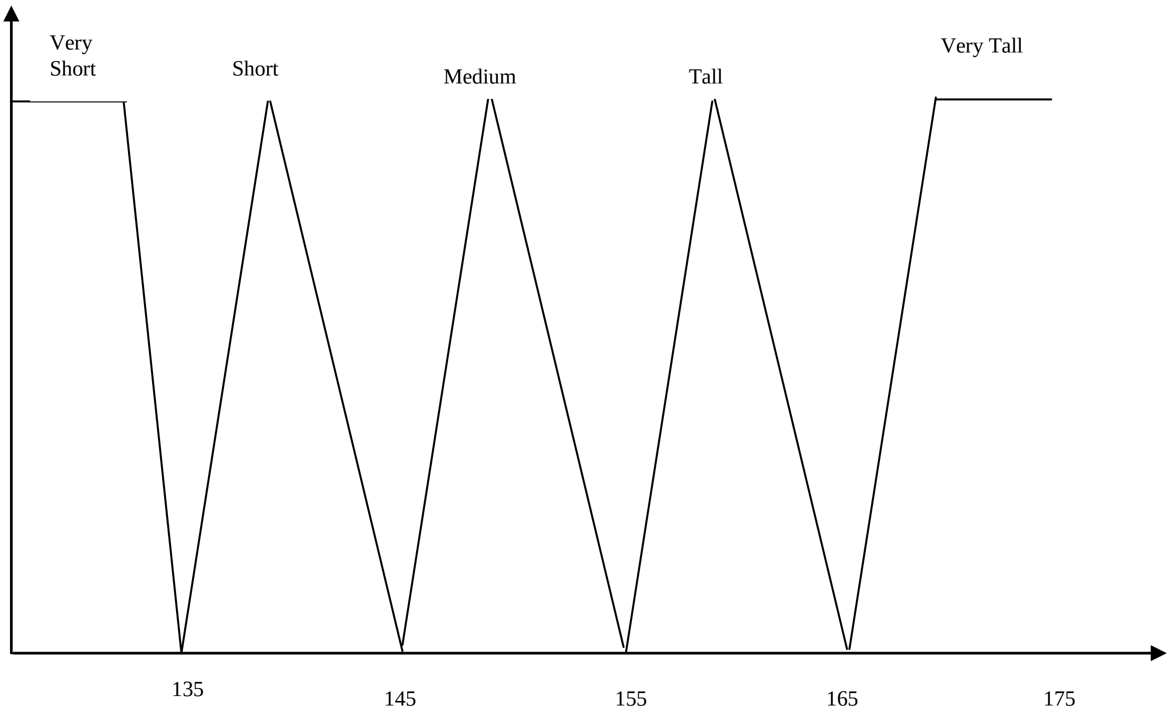

Figure 1. Membership functions representing the linguistic terms Very Short, Short, Medium, Tall and Very Tall

3.2 Fuzzy Logic Based Constraints

Definition 1: A Constraint C_s on a set of variables $S = \{x_1, x_2, ...x_r\}$ is a pair $C_s = (S,R_s)$, where R_s is a subset on the product of these variables' domains:

$$R_s \subseteq D_1 \times \cdots \times D_r \rightarrow \{0,1\}$$

The number *r* of variables a constraint is defined upon is called parity of the constraint. The simplest type is the *unary constraint*, which restricts the value of a single variable. Of special interest are the constraints of parity two, called *binary constraints*. A constraint that is defined on more than two variables is called a *global constraint*.

A constraint is usually defined as a pair consisting of a set of variables and a relation on these variables. This definition gives us the availability to model different types of uncertainty in schema matching. Uncertainty in semantic mappings between data sources can be modeled by exploiting fuzzy relations while other sources of uncertainty can be modeled by making the variable set a fuzzy set.

Definition 2. A Fuzzy Constraint C_μ on a set of variables $S = \{x_1, x_2, ...x_r\}$ is a pair $C_\mu = (S, R_\mu)$, where the fuzzy relation R_μ, defined by μ_R

$$: \coprod_{x_r \in \mathrm{var}(C)}. \qquad D_i \rightarrow [0,1] \text{ where}$$

where μ_R is the membership function indicating to what extent a tuple v satisfies $C_{\mu..}$ μ_R (v) = 1 means v totaly satisfies $C_{\mu..}$ μ_R (v) = 0 means v totally violates C_μ, while $0 < \mu_R (v) < 1$ means v partially satisfies $C_{\mu.}$

Definition 3: A Fuzzy Constraint optimization Q_μ is a 4-tuple $Q_\mu = (X,D, C_\mu,g)$ where X is a list of variables, D is a list of domains of possible values for the variables, C_μ is a list of Fuzzy Constraints each of them referring to some of the given variables, and g is an objective function to be optimized.

4 A Framework of XML Schema with Fuzzy Constraints

In this paper, we have introduced a fuzzy constraint-based framework for XML schema. Our approach is able to handle uncertainty in schema matching with XML document by exploiting fuzzy constraints. Moreover, our framework is generic which has the feature to deal with different schema representations. We would like to provide Fuzzy Constraints in XML Schema. In this sense, we must allow the specification of user defined linguistic terms in XML Schema. In our fuzzy logic based extension, the allowed linguistic terms are predicates. The general schema for fuzzy term specification is given in Listing 1. We specify fuzzy terms through XML using the XML Schema language.

Listing 1. Xml Schema for Fuzzy Terms

```
<xsd:complexType name="xsd:term">
<xsd:all>
<xsd:group ref="xsd:predicate"/>
</xsd:all>
</xsd:complexType>
```

The XML Schema for the fuzzy predicate definition is shown in Listing 2. A fuzzy predicate is specified with a name, a domain and the membership function of the fuzzy set defining the predicate. The membership function that may be of two kinds: trapezium and triangle.

Listing 2. Xml Schema for Fuzzy Predicate

```
<xsd:group name="xsd:predicate">
<xsd:sequence>
<xsd:attribute name="name"
type="xsd:ID"/>
<xsd:element name="domain"
type="xsd:string"/>
<xsd:choice>
<xsd:group ref="xsd:trapezium"/>
<xsd:group ref="xsd:triangle" />
</xsd:choice>
</xsd:sequence>
</xsd:group>
```

A trapezium shape membership function is described by the sequence of its four inflection points x values: x1, x2, x3 and x4. The specification of membership function in XML Schema is given in Listing 3. The semantics is as follows: The first is where the function begins to growth form zero to one, the second is the first point in which the function raise the degree one, between these two points the satisfaction degree increases in straight line, between second and third point, the function gives the constant satisfaction degree one, the fourth point is where the element becomes completely excluded, between the third and the fourth, the satisfaction degree decreases proportionally.

Listing 3. Xml Schema for Trapezium Shape Membership Function

```
<xsd:group name="xsd:trapezium">
<xsd:sequence>
<xsd:element name="x1"
type="xsd:decimal"/>
<xsd:element name="x2"
type="xsd:decimal"/>
<xsd:element name="x3"
type="xsd:decimal"/>
<xsd:element name="x4"
type="xsd:decimal"/>
</xsd:sequence>
</xsd:group>
```

A triangle shape membership function is described by the sequence of its three inflection points x values: x1, x2, and x3. The specification of membership function in XML Schema is given in Listing 4.

Listing 4. Xml Schema for Triangular Membership Function

```
<xsd:group name="xsd:triangle">
<xsd:sequence>
<xsd:element name="x1"
type="xsd:decimal"/>
<xsd:element name="x2"
type="xsd:decimal"/>
<xsd:element name="x3"
type="xsd:decimal"/>
</xsd:sequence>
</xsd:group>
```

The Listing 5 gives an example of user defined fuzzy predicate with trapezium shape membership function.

Listing 5. Xml Document Example for Fuzzy Predicates

```
<predicate name = "high temperature">
<heat unit> Celsius</heat unit>
<trapezium> <x1> 40 </x1> <x2> 45
</x2>
<x3> 50 </x3><x4> 55 </x4>
</trapezium>
</predicate>
```

5 Examples for Fuzzy Constraints with Xml Schema

In this section, we give a simple example of an XML Schema with the data of an employee such as Empid, name, salary, height and designation. We implement Fuzzy Constraint for the Height attribute of the employee. In the XML Schema, we group the employees according to Fuzzy Linguistic term based constraints Very Short, Short, Medium, Tall and Very Tall. All the linguistic terms are associated with triangular membership functions. The Linguistic term based membership functions are defined in the formulae H_1, H_2, H_3, H_4, H_5. The Linguistic term based triangular membership functions are illustrated in the Figure 1. If the user violates the range of these membership functions, the XML Schema will prompt error messages. Listing 6 defines how the fuzzy constraints are implemented with XML Schema for Employee XML Document. The validated Employee XML document against the XML Schema with Fuzzy Constraints is shown in the Listing 7. Domain Integrity Constraints will be implemented using Fuzzy rules. We can enforce intelligent business rules using Triangular or Trapezoidal fuzzy membership functions.

Listing 6. XML Schema with Fuzzy Constraints for Employee XML Document

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema
xmlns:xs="http://www.w3.org/2001/XMLSchem
a" elementFormDefault="qualified"
attributeFormDefault="unqualified">
        <xs:element name="employees">
        <xs:complexType>
        <xs:choice maxOccurs="unbounded">
        <xs:element
name="VeryTallEmployees">
        <xs:complexType><xs:sequence>
        <xs:group ref="target:emp"/>
        <xs:group
ref="target:VeryTall"/></xs:sequence></xs:com
plexType>
        </xs:element>
        </xs:choice></xs:complexType></xs:el
ement>

        <xs:group name="emp"><xs:sequence>
        <xs:element name="empid"
        type="string" />
        <xs:element name="ename"
type="string" />
        <xs:element name="salary"
type="decimal"/>
        <xs:element name="designation"
type="string"/>
        </xs:sequence> </xs:group>

        <xs:group name="Very Tall">
        <xs:sequence>
        <xs:element name="height"
type="integer" />
        <xs:assert test="((185-@height)/10 > 0)
and ((185-@height)/10 <1)">
        <xs:annotation>
        <xs:appinfo>Fuzzy Very Tall constraint
        is Violated. Height should be between
        175 and 185     </xs:appinfo>
        <xs:documentation>
         When this assertion fails, the content of
the above "appinfo" is used
        to produce the schema error message.
        </xs:documentation>
        </xs:annotation></xs:assert>
        </xs:sequence></xs:group>
```

Listing 7. The Validated Employee XML Document

```
<?xml version="1.0" encoding="UTF-8"?>
<employees
xmlns:xsi="http://www.w3.org/2001/XMLSche
ma-instance"
xsi:noNamespaceSchemaLocation="D:\2NDSIX
~1\Fuzzy_Schema.xsd">
<emp><VeryTallEmployees>
<empid>B2000</empid>
<ename>Jeff</ename>
<salary>8000</salary>
<designation>Software Engineer</designation>
<height>181</height>
</VeryTallEmployees></emp></employees>
```

6 Conclusion

In this paper, we proposed a general XML Schema definition for representing fuzzy information in XML documents. We have introduced a fuzzy constraint-based framework to validate the XML document against the XML schema. We introduced Trapezoidal and Triangular fuzzy membership functions for implementing Fuzzy Constraints. We have defined the mechanism to represent fuzzy data along with crisp data in XML Schema .Our approach is able to handle uncertainty in schema matching by exploiting fuzzy constraints. In this paper, we have focused only on Domain Integrity constraints using Fuzzy Logic. Because this is an important control for XML based applications for implementing data consistency. We have to do further research on Entity Integrity Constraints and Referential Integrity Constraints for Native XML Databases. In future, we plan to do further research on introducing Fuzzy Triggers on the basis of XML Schema with Fuzzy Constraints.

References

[1] Buckles B. P. and Petry F. E., "A fuzzy representation of data for relational databases" ,*Fuzzy Sets and Systems*, Vol.7, pp.213-226, 1982.

[2] Barbara Oliboni, Gabriele Pozzani, "An XML Schema for Managing Fuzzy Documents" Technical Report, University of Verona, Italy, 2008

[3] Chaudhri, A., Rachid, A., Zicari, R. XML Data Management. Addison-Wesley,
1st Edition. 2003.

[4] Yunsheng Liu, Hao Zhong, Yi Wang, Capturing XML Constraints with Relational Schema, in Proceedings of the Fourth IEEE International Conference on Computer and Information Technology, 2004

[5] Abhishek Gaurav, Reda Alhajj, "Incorporating Fuzziness in XML and Mapping Fuzzy
Relational Data into Fuzzy XML", in Proceedings of the ACM Symposium on Applied Computing , France, April, 2006.

[6] Alsayed Algergawy, Eike Schallehn, and Gunter Saake, Fuzzy Constraint-based Schema Matching Formulation, Scientific International Journal for Parallel and Distributed Computing, Vol 9, No 4, 2008.

[7] Carlo Combi, Barbara Oliboni, and Rosalba Rossato, Evaluating Fuzzy Association Rules on XML Documents, in proceedings of International Conference on Computational Intelligence, Theory and Applications, 2004,Germany.

[8] Klaus Turowski,, Uwe Weng, "Representing and processing fuzzy information — an XML-based approach", Journal of Knowledge Based Systems, Vol 15,January 2002, pp 67-75.

[9] Zadeh L. A., "Fuzzy Sets," Information and Control, Vol.8,Academic Press, New York, pp.338-353, 1965.