

Generating Input Documents for Testing XML Queries with ToXgene

Dae S. Kim-Park, Claudio de la Riva, Javier Tuya, José García-Fanjul
Computing Department, University of Oviedo
Campus of Viesques, s/n 33204 – Gijón (SPAIN)

kim_park@lsi.uniovi.es, claudio@uniovi.es, tuya@uniovi.es, jgfanjul@uniovi.es

Abstract

Modern software technologies use the XML language as a preferred means of data interchange and representation, which can be accessed using XML queries. Testing these queries can be difficult because they use complex data as input in form of XML documents. This work addresses this problem by defining a query-aware procedure for design and generation of test input data for XML queries. The test process is based on the category-partition method and permits the identification of categories and choices in the XML query and schema in order to obtain test case specifications. These test case specifications are mapped to the input language of the ToXgene data generator and then XML documents are generated by running it.

1. Introduction

Test generation is typically manual and thus tedious and error-prone when generating complex or a large number of tests. Manual generation of test input values is particularly hard for code that takes complex data structures as input. This is the case for software applications that manipulate XML documents, as for example Web and Data Services.

During recent years, there has been an important body of research on automatic generation of XML data for web-based applications and on development of XML data generators. Among them, the ToXgene tool [4] represents the state of the art of the current XML data generators, because it provides a flexible and declarative way of generating synthetic XML data. However, the specific practices and tools in generating such test input values frequently ignore the XML queries used for manipulating data, because they generate XML data independent of the query.

This abstract is concerned with the systematic and automatic test input generation for XML queries. We

propose a query-aware technique based on the Category-Partition Method [3] for designing test input cases and provide the automatic support for generating test input cases (XML documents) with the ToXgene XML data generator.

2. Method Overview

The method, which is depicted in Figure 1, comprises two main steps. Step 1: Obtaining test case specifications (test frames) and Step 2: Generating test documents for the XML query.

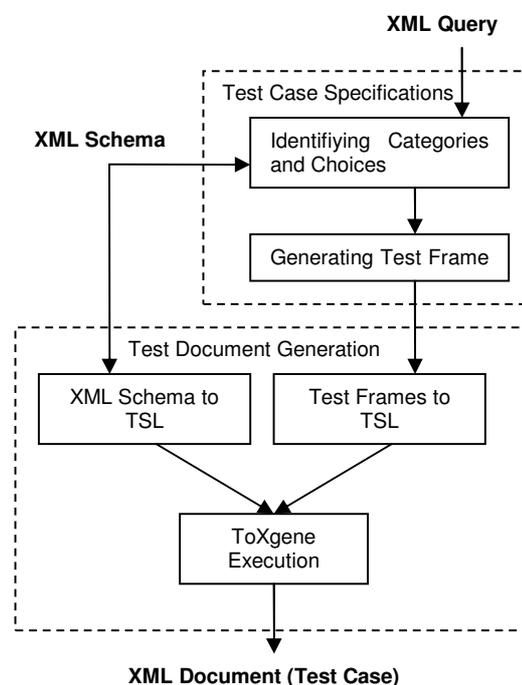


Figure 1. Overview of the method

2.1 Test Case Specifications

The Category Partition Method (CPM) [3] consists in the partition of data input into a set of categories that identifies the main characteristics of the input parameters. Each category is then divided into a set of choices that represent the different values which they can take in the category.

In a previous work [2], we presented a partition-based approach for testing XML queries. Categories and choices are identified by systematically examining both the XML schema and the XML query. Three types of choices are determined:

- Existence and Empty choices: They represent the existence and empty value of an element in the query.
- Number-of-elements choices: They denote the number of occurrences of an element in the query.
- Expression choices: They are related to the predicates with a constant value and the predicates with correlated content.

The choices are then combined in each category to form test frames (test case specifications). More details can be found in [2].

2.2 Test Document Generation

ToXGene [1] is a flexible, declarative and template-based generator for XML data mainly used for benchmarking purposes. The data generation is guided by templates that specify a schema for the desired data as well as rules (*genes*) for populating such a schema. The templates are XML documents represented using the notation of TSL (Template Specification Language) augmented with annotations for guiding the data generation process.

Our approach consists of the following steps: 1) transform the XML Schema to the TSL template, 2) transform each choice of the test frame to the TSL template, and 3) execute ToXgene using the TSL template as input.

XML Schema representation. Given an XML schema, the transformed TSL template consists of the declaration of a set of genes for describing the schema content. The transformation is done by mapping the XML simple and complex types to the TSL simple and complex types in a straightforward way.

Coding the test frames. Even though ToXgene has helpful capabilities for synthetic XML data generation, the TSL template of the schema representation is *query-unaware*. For the purpose of breaking the isolation between the query and the TSL schema

representation, we also code the generated test frames in the TSL template.

The different types of choices in a test frame are included in the template expanding the structural declaration of elements with extra constraints in the content genes. Existence and Empty choices as well as Number-of-Elements choices are represented in TSL by defining particular boundaries for each implied element. For the expression choices, constant value conditions are specified as TSL value restrictions. Finally, coding choices with correlated content has a more complex approach that involves the application of some coding patterns to declare TSL lists. The lists contain all the data that takes part in the element relations.

3. Conclusions and Future Work

This research abstract summarizes a technique which generates test input values for XML queries. It is based on the well-known Category Partition Method used in the software testing community and uses ToXgene as the XML document generator by augmenting the structural definition of XML data with features related to the XML query. Such procedure enables the automatic generation of test input values addressed to cover important features of the XML query.

The immediate line of work is addressed to reducing the number of generated test documents.

Acknowledgements

This work was partially funded by the Department of Education and Science (Spain) and ERDF funds within the National Program for Research, Development and Innovation, project Test4SOA (TIN2007-67843-C06-01) and the RePRIS Software Testing Network (TIN2007-30391-E). Dae S. Kim-Park is also supported by Department of Education and Science (Spain).

References

- [1] Barbosa, D., Mendelzon, A.O. Declarative Generation of Synthetic XML Data. *Software Practice and Experience*, vol 36, pp. 1051-1079, 2006
- [2] De la Riva, C., García-Fanjul, J., Tuya, J. A partition-based approach for XPath Testing. In *Proceedings of International Conference on Software Engineering Advances*, pp. 17-23, 2006.
- [3] Ostrand, T.J., Balcer, M.J. The Category-Partition Method for specifying and generating functional tests. *Communications of the ACM*, vol. 31. no. 6, pp. 676-686, 1988.
- [4] ToXgene. <http://www.cs.toronto.edu/tox/toxgene/>, 2005