ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Systems@ETH Zürich

Master's Thesis Nr. 327

Systems Group, Department of Computer Science, ETH Zurich

A Test Suite for Rumble

by

Stevan Mihajlovic

Supervised by

Dr. Ghislain Fourny, Prof. Dr. Gustavo Alonso

October 1, 2020 - April 1, 2021

D INFK

Abstract

Insert some short text here

Acknowledgements

Write some sad story and then a positive story. Don't forget to mention all the people

Contents

Chapter 1

Introduction

1.1 Motivation

Mention Rumble. Mention JSONiq. Mention XQuery test suite. Mention similarity and reusage. Mention Goals.

The increasing amount of data available to process, as well as the ever-growing discrepancy between storage capacity, throughput and latency, has forced the database community to come up with new querying paradigms in the last two decades. Data became nested and heterogeneous (JSON), and is increasingly processed in parallel (Spark). In order to make querying more efficient and accessible, Rumble [MFI+20] is an engine that automatically runs queries on semi-structured and unstructured documents on top of Spark, using the JSONiq language.

JSONiq [jso11]is a functional and declarative language that addresses these problems with its most useful FLWOR expression which is the more flexible counterpart of SQL's SELECT FROM WHERE. It inherits 95% of its features from XQuery, a W3C standard.

The XQuery/XPath 3.* Test Suite (QT3TS) [W3C94] provides a set of tests with over 30000 test cases designed to demonstrate the interoperability of W3C XML Query Language, version 3.0 and W3C XML Path Language implementations.

The high level idea of this work is to implement a Test Driver that can directly use QT3TS in order to test and verify Rumble implementation.

1.2 Related Work

Another section

1.3 Contributions

Another section

1.4 Outline

Another section

Chapter 2

XQuery/XPath 3.* Test Suite(QT3TS)

2.1 Analysis

We need to show how Saxon implementation is verified using the test suite. The we need to show the class that we want to reproduce. We need to mention that everything is XML and that we will reuse Saxon for parsing the XML, backwards compatibility with everything except XSLT

In this chapter, we will discuss design decisions that we have made during the development of Test Driver. The idea is to develop Test Driver completely independently from Rumble by maintaining the code outside of Rumble. We view Rumble as black-box and the single point of communication with Rumble should only be via the Rumble Java public API. Therefore, we have decided to implement Test Driver as Java Console Application. Furthermore, as Rumble is also written in Java, it is probably the best option in case that not everything is possible to be achieved by treating Rumble as the black-box.

The XQuery/XPath 3.* Test Suite (QT3TS) is publicly available at W3C Public CVS Repository under module name 2011/QT3-test-suite [W3C11]. Since April 1st 2019, CVS tree has been discontinued and the repository has been migrated to W3C Public GitHub repository [W3C20]. The tests are published as a set of files - test sets containing in total more than 30000 test cases. The tests are published as a set of files, mostly in XML format. W3C does not supply a Test Driver for executing the tests. Instead, for each implementation a Test Driver should be written. As these test sets are mostly written in XML format, the first component that our Test Driver will require is the XML parser.

For parsing XML, we have decided to use Saxon [Kay20] - Open Source XSLT & XQuery processor developed by Saxonica Limited. First of all, Saxon itself is one of the implementations for which Test Driver was also implemented. Based on Results Report [Kay16], it passes more than 99,9% of the QT3TS test. Furthermore, Saxons implementation of the Test Driver can be used as a baseline for developing our own Test Driver.

2.2 Phases

2.3 JSONiq

The JSONiq extension to XQuery allows processing XML and JSON natively and with a single language. This extension is based on the same data model as the core JSONiq and is based on the same logical concepts. Because of the complexity of the XQuery grammar, the JSONiq extension to XQuery has a less pleasant syntax that the JSONiq core. When designing the Test Converter, we could have decided to use either XML or JSON as the underlaying language. However, as our Test Driver was already implemented in the previous phase and was expecting XML as input and using the before mentioned Saxon for parsing it, we have decided to keep the same language for output of the Test Converter.

Chapter 3

Concluding Remarks

3.1 Overall Summary

sum up the results

3.2 Open Problems

further research here

Bibliography

[jso11] jsoniq.org. Jsoniq - the json query language, 2011.

[Kay16] Michael H. Kay. Qt3 test suite result summary, 2016.

[Kay20] Michael H. Kay. Saxon: The xslt and xquery processor, 2020.

[MFI+20] Ingo Müller, Ghislain Fourny, Stefan Irimescu, Can Berker Cikis, and Gustavo Alonso. Rumble: Data independence for large messy data sets, 2020.

[W3C94] W3C. Xquery/xpath/xslt 3.* test suite, 1994.

[W3C11] W3C. Xquery/xpath/xslt 3.* test suite cvs repository, 2011.

[W3C20] W3C. Xquery/xpath/xslt 3.* test suite github repository, 2020.

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

A Test Suite for Rumble

Authored by (in block letters):
For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Mihajlovic	Stevan

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zürich, October 1, 2020	Stevan Mihajlovic

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.