MASARYK UNIVERSITY

FACULTY OF INFORMATICS

Implementation of provenance chains traversal

Bachelor's Thesis

TOMÁŠ ZOBAČ

Brno, Fall 2023

MASARYK UNIVERSITY

FACULTY OF INFORMATICS

Implementation of provenance chains traversal

Bachelor's Thesis

TOMÁŠ ZOBAČ

Advisor: RNDr. Rudolf Wittner

Department of Computer Systems and Communications

Brno, Fall 2023


FACULTAS ARTIS INFORMATICÆ
UNIVERSITAS MASARYKIANA

Declaration

Hereby I declare that this paper is my original authorial work, which I have worked out on my own. All sources, references, and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

Tomáš Zobač

Advisor: RNDr. Rudolf Wittner

Acknowledgements

TODO

Abstract

Provenance is a standardized information type that documents the history of an object. It can hold information, such as the origin of an object or previous actions performed on it. This information can be serialized into one of the many supported file formats (e.g., PROVN, XML). These files can then be interconnected, resulting in the creation of a provenance chain. This thesis aims to implement a library for traversing said provenance chains, retrieving information about the precursors or successors of an entity represented in one of the files in the current chain, and optionally retrieving the type of actions performed on the object by the retrieved precursors/successors. The implementation will simulate the operational environment by providing a command-line user interface from where the user can call the mentioned actions on a set of prefactured simulation files. The implementation will follow a W3C PROV-DM standard for provenance notation, and the simulation files will be of a PROVN serialization of provenance data. Additionally to the traversing operations, it will also implement the generation of a provenance metadata to make the simulation of a traversal possible.

Keywords

Java, provenance, SOBHA, ÚVT

Contents

Introduction **1**

1 Provenance **2**

2 PROV-DM **3**

3 Used technologies **4**
3.1 Java 4
3.2 Maven 4
3.3 Jetbrains IDEA 4
3.4 ProvToolBox 4
3.4.1 PROV MODEL 4
3.4.2 PROV INTEROP: LIGHT 5
3.5 Jackson Databind 5
3.6 GitLab4J API 5
3.7 JLine Bundle 5
3.8 Jansi 5
3.9 Apache Maven Shade Plugin 6

4 Implementation **7**
4.1 Components 7
4.2 Simulation 7
4.3 Algorithm 8
4.4 Deployment 9

5 Manual **10**
5.1 Set-up 10
5.1.1 Clone the repo 10
5.1.2 Init simulation files 10
5.2 Building 11
5.3 Omitting the simulated environment 12

6 Real life application **13**

Bibliography **15**

List of Tables

List of Figures

Introduction

Theses are rumoured to be “the capstones of education”, so I decided to write one of my own. If all goes well, I will soon have a diploma under my belt. Wish me luck!

1 Provenance

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

2 PROV-DM

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

3 Used technologies

3.1 Java

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.2 Maven

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.3 Jetbrains IDEA

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.4 ProvToolBox

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.4.1 PROV MODEL

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.4.2 PROV INTEROP: LIGHT

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.5 Jackson Databind

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.6 GitLab4J API

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.7 JLine Bundle

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.8 Jansi

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

3.9 Apache Maven Shade Plugin

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies.

4 Implementation

4.1 Components

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

4.2 Simulation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci

augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

4.3 Algorithm

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

4.4 Deployment

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

5 Manual

5.1 Set-up

In order to use the library, correctly setup Maven and Java are required. For user convenience the implementation currently lives in multiple git repositories (two on FI and ICS GitLab sites and one on my personal GitHub page).

To download the library, you can either clone the repository or use the download button to get the whole repository bundled to zip or other archive

5.1.1 Clone the repo

1. Clone/Code
2. Clone with HTTPS
3. In the console, run `git clone <copied_url>`

5.1.2 Init simulation files

In order for the simulated environment to run, you need to pull the required files

1. Open the cloned repo in the console
2. Move to the submodule using

   ```
   cd.\src\main\resources\bthesis-provenancechain-digpat
   ```

3. Run **`git submodule foreach git fetch -tags`** (There should be no output)
4. Finish by running `git submodule update --init --recursive`
5. The `bthesis-provenancechain-digpat` submodule should now be populated with `.provn` files

5.2 Building

The implementation uses a Maven Shade plugin for jar packaging

1. Open the cloned repo in the console
2. Run ‘mvn clean package‘
3. Execute the created jar with ‘java -jar .
   target
   BThesis-ProvenanceChain-VERSION-shaded.jar‘
   - alternatively, if the targeted environment doesn’t have a JRE, you can create an exe installer using Java’s prepackaged ‘jpackage‘ command line tool

5.3 Omitting the simulated environment

In order to simulate the traversal, the implementation uses multiple classes and files to provide the required objects for the traversing algorithm to work. For clarity, these classes are moved to packages ‘bthesis.provenancechain.simulation‘ and ‘bthesis.metageneration‘, while the required files are located in the submodule ‘.
src
main
resources
bthesis-provenancechain-digpat‘ mentioned before. All of these can be removed as long as the classes needed are sufficiently replaced.

6 Real life application

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer finibus commodo leo. Nullam blandit imperdiet magna, sit amet tempor tortor sagittis vitae. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Cras elementum diam vel eros tristique, at maximus tortor ultricies. Curabitur urna magna, dictum at porta rutrum, congue nec justo. Nam ac rhoncus lectus. Ut feugiat volutpat ornare. Mauris quis neque nec lorem vestibulum iaculis. Proin posuere nisi eget nisi tristique, eu tempus nibh ultricies.

Nunc sagittis id purus quis facilisis. Etiam ullamcorper massa ut justo volutpat, quis malesuada lectus fermentum. Maecenas orci augue, dictum vel risus ut, finibus tincidunt magna. Praesent semper vehicula nulla, eu dignissim elit aliquet in. Mauris ultricies tortor sed justo congue, eu pretium sapien egestas. Aliquam erat volutpat. Duis elementum urna nec felis pharetra, egestas aliquam turpis aliquam. Donec commodo, felis vitae pulvinar interdum, nisl felis feugiat est, vel dictum augue lacus nec elit. Etiam a cursus ligula. Sed luctus nisl at rhoncus ullamcorper. Suspendisse vitae sapien porttitor, auctor quam ac, volutpat tellus. Phasellus cursus, enim vel tristique porttitor, velit dolor viverra sapien, eu lacinia elit erat cursus augue. Fusce sodales venenatis lacus, eget rutrum lorem bibendum vel.

After linking a bibliography database files to the document using the `\thesissetup{bib={file1,file2,...}}` command, you can start citing the entries. This is just dummy text (1) lightly sprinkled with citations (2, p. 123). Several sources can be cited at once: [1, 2, 3]. "Camel drivers and gatecrashers" was written by Greenberg in 1998. We can also produce Greenberg [2] or (Greenberg [2], 1998). The full bibliographic citation is: *GREENBERG, David. Camel drivers and gatecrashers: quality control in the digital research library. In: HAWKINS, B.L; BATTIN, P (eds.).* The mirage of continuity: reconfiguring academic information resources for the 21st century. *Washington (D.C.): Council on Library and Information Resources; Association of American Universities, 1998, pp. 105–116*. We can easily insert a bibliographic citation into the footnote[1].

1. GREENBERG, David. Camel drivers and gatecrashers: quality control in the digital research library. In: HAWKINS, B.L; BATTIN, P (eds.). *The mirage of continuity:*

Bibliography

1. BORGMAN, Christine L. *From Gutenberg to the global information infrastructure: access to information in the networked world*. 1st ed. Cambridge (Mass): The MIT Press, 2003. ISBN 0-262-52345-0.
2. GREENBERG, David. Camel drivers and gatecrashers: quality control in the digital research library. In: HAWKINS, B.L; BATTIN, P (eds.). *The mirage of continuity: reconfiguring academic information resources for the 21st century*. Washington (D.C.): Council on Library and Information Resources; Association of American Universities, 1998, pp. 105–116.
3. THÀNH, Hàn Thé. *Micro-typographic extensions to the TEX typesetting system* [online]. Brno, 2001 [visited on 2016-12-09]. Available from: `http://www.pragma-ade.nl/pdftex/thesis.pdf`. PhD thesis. The Faculty of Informatics, Masaryk University.

---

reconfiguring academic information resources for the 21st century. Washington (D.C.): Council on Library and Information Resources; Association of American Universities, 1998, pp. 105–116.