Outline of the Thesis

Table of Contents

1 Introduction

1.3 Report outline

1.2 Project objectives

2 Context survey

2.1 Defining the main terms and concepts

2.1.1 GAP

2.1.2 Jupyter Notebooks

2.1.3 Citation

2.2 Related work

2.2.1 FORCE11

2.2.2 ArXiv

2.2.3 GAP website

2.2.4 Dr Konovalov

2.2.5 Software Citation Implementation Challenges by D. Katz and VA

2.4 Remaining / Research questions

3 Ethics

2.5 Software Requirements specification

2.6 Software engineering process outline

5 Design

5.2 Design and Implementation of the data scraping notebook

5.3 Design and Implementation of the data analysis notebook

6 Analysis

6.1 Evaluation and critical appraisal

7 Conclusions

Appendices

6.2 Testing summary

User Manual

Other appendices

9 References

2 Context survey and literature review

In this project my main goal will be analysing the GAP Citations dataset to gain insights, make conclusions, follow various trends and anomalies. My second goal however will be to research some of the most prominent related works/papers in the field of Software citation, produce a refined set of recommendations for best practices and a unified software citation standard and finally evaluate how well do the GAP citations documented so far adhere to the best software citation principles.

2.1 Defining the main terms and concepts

GAP (Groups, Algorithms and Programming) is a mathematical software system for discrete computational algebra with particular emphasis on computational group theory. GAP offers a range of vast libraries containing functions, algebraic objects and algebraic algorithms written in the GAP programming language. It is important to note that GAP is an open-source and freely distributed system, which can be modified to suit specific needs, or used as a base to build on, it can be downloaded from the official website <https://www.gap-system.org/> . Nevertheless, if a published work uses parts of GAP, it needs to be cited just like scientific journals or books are. [1]

GAP was awarded the *ACM/SIGSAM Richard Dimick Jenks Memorial Prize for Excellence in Software Engineering applied to Computer Algebra* in July 2008. [1]

Along with Aachen, Braunschweig, Fort Collins and Kaiserslautern, The University of St Andrews is one of the 5 major GAP centres, by which the maintenance and further development effort are coordinated. [1]

JNs

Citations, also called references, uniquely identify a source of information used in published works such as books and articles. [2] Citing is important firstly because credit must be given where it is due – if we use others’ ideas, texts, findings, research or software in our work then we must acknowledge all of our sources. The second main reason for citing to be crucial to the scientific world is credibility – if a work is based on one or more highly respected and acknowledged sources then this work will undoubtedly be more credible and will have more weight.

2.2 Related work

The general trend amongst scientific communities today is that software is a legitimate form of intellectual property, just like books and articles and as such it should indeed be cited and referenced consistently. Unfortunately, at this point in time there is still no unified, globally approved standard for software citation. However, many credible and respected institutions and organisations are working to develop and start implementing sets of rules and standards, which will continuously evolve and increase their scope. Eventually the best sets of principles will be joined to form global standards for citing software and other digital intellectual property.

2.2.1 FORCE11

According to the *Software Citation Principles* document by the FORCE11 Working Group and various authors, in recent years, the processes of studying and research are becoming more and more digital. The results of research are no longer simply articles and books but are now extended to include software and vast amounts of electronically stored data. Other digital elements such as graphic presentations (posters, slides), interactive visualisations (charts, graphs, maps), websites (blogs and forums) and multimedia (audio and video) have become inseparable part of contemporary education and research. The fruits of intensive scholarly labor in the form of knowledge and insights are interwoven in all these digital elements today, just as they were contained only in books and articles in the past. Therefore, the principles for citation, credit and referencing which apply for papers need to be updated, developed and significantly extended to start encompassing the myriad of all these new and interlinked digital resources. [3]

2.2.2 Research Software Sustainability and Citation

This document is a collaboration between Stephan Druskat, Daniel S. Katz, and Ilian T. Todorov, who are based in Germany, USA and UK respectively, here the authors point out in a concise manner the main contributions of software citation to software sustainability and also list the major challenges standing in the way of software citation.

2.2.3 GAP website

2.2.4 Dr Konovalov slides

2.2.5 Software Citation Implementation Challenges by D. Katz and VA

Workflow Time Plan

References: