

SET09103 Advanced Web Technologies

Coursework Part #1

Vinh Phat Tu
40507973

October 2023

Contents

1	Introduction	2
2	Data Preparation	2
2.1	Data Cleaning	2
2.1.1	Missing header row	2
2.1.2	OpenRefine	2
2.2	Data Transformation	2
2.3	Data Framework and Visualisation	2
2.4	Scientific Analysis	2

1 Introduction

The aim of this coursework is to prepare and analyse a data set using tools such as OpenRefine and Weka and languages such as Python.

2 Data Preparation

2.1 Data Cleaning

2.1.1 Missing header row

The raw data set had no header row and thus a row was manually added with the respective attribute names in Microsoft Excel.

After investigating, it was found that OpenRefine does not offer a simple way to insert a header row and thus Excel was used for insertion.

2.1.2 OpenRefine

OpenRefine offers multiple ways of inspecting and cleaning large amount of data through common transformations,

2.2 Data Transformation

2.3 Data Framework and Visualisation

2.4 Scientific Analysis

Here are some sample citations. An article by Einstein [1] and a book by Dirac [2].

References

- [1] Albert Einstein. “Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies]”. In: *Annalen der Physik* 322.10 (1905), pp. 891–921. DOI: <http://dx.doi.org/10.1002/andp.19053221004>.
- [2] Paul Adrien Maurice Dirac. *The Principles of Quantum Mechanics*. International series of monographs on physics. Clarendon Press, 1981. ISBN: 9780198520115.