An Infrastructure to Store and Analyse Seismic Data as Suffix Trees

MSc Data Analytics - Project Proposal

Tom Taylor

April 2016

This report is substantially the result of my own work except where explicitly indicated in the text. I give my permission for it to be submitted to the JISC Plagiarism Detection Service. I have read and understood the sections on plagiarism in the Programme Handbook and the College website.

Contents

1 Abstract

The purpose of this project is to develop an infrastructure and tool set for converting raw seismic time series data in to a searchable string and then to store this data as a suffix tree for fast searching and analysis. An interface would then be developed to enable the searching of these suffix trees and provide visualisation of the data.

2 Background

Seismic waves are recorded as movement over three axis: vertical (hereafter referred to as **z**) alongside horizontal in terms of north-south(**n**) and east-west(**e**).

Some words about SAX.

Some words about Suffix Trees.

3 Objectives

1. Interpretation of raw data from seismic stations and storage as time-series data in a time-series database such as OpenTSDB for easy access during conversion to SAX and for later rendering during interactions with the data.
2. Conversion of the data to SAX and storage in a suffix tree.
3. An interface for searching and viewing the raw data.

4 Approach

5 Plan