

Requirement Document: Repository and Mining of Temporal Data

Team Members:

Jessica Nguy jnguy2014@my.fit.edu
Siomara Nieves snieves2014@my.fit.edu

Faculty Sponsor and Client:

Philip Chan pkc@fit.edu

Table of Contents

1. Introduction
 - 1.1. Purpose / Scope
 - 1.2. Definitions
 - 1.3. Overview
2. Specific Requirements
 - 2.1. uploadData
 - 2.2. questionOne
 - 2.3. questionTwo
 - 2.4. questionThree
 - 2.5. setTag
 - 2.6. setTarget
 - 2.7. narrowData
 - 2.8. visualizeData
 - 2.9. metadata
 - 2.10. webApp
 - 2.11. exportCSV

1. Introduction

1.1. Purpose / Scope

This document describes requirements and specifications for RepoMining, which stands for Repository and Mining of Temporal Data. RepoMining is a web-based repository of temporal data. The system aims to integrate data from different data sets and different repositories into one centralized site, allow for users to upload and download data, answer user queries on specific data values, visualize the data in graphical form, and feature a simplified GUI for data providers and data customers to use.

The system is designed to answer 3 questions:

1. Is there a significant change in the target variable?
2. Why was there a significant change, and what are the top variables that are correlated with the target variable?
3. What is the value in the next timestamp?

1.2. Definitions

GUI - Graphical User Interface. It is what the user sees when using the program.

CSV - Comma Separated Value. It is the data format accepted by the program.

Repository - A central location where data is managed.

Data Providers - Users who upload data to the site via a .CSV file. Values in data possess timestamps either in vertical or horizontal format, and the data must be numerical.

Data Consumers - Users who can use the data on the site to answer questions. A data consumer can be a data provider also. Data customers must specify a target variable.

Q1 - Shorthand for “Is there a significant change in the target variable?”

Q2 - Shorthand for “Why was there a significant change, and what are the top variables that are correlated with the target variable?”

Q3 - Shorthand for “What is the value of the target variable in the next timestamp?”

Meta-data - Descriptor that contains the description of each variable, start/end timestamps, tags, and a public/private setting.

Correct Format - csv file that has timestamps going across the first row and variables listed in the first column, or timestamps going down in a column and variables listed in the first row.

1.3. Overview

Data is located everywhere, but in different formats, from different sources, and not easy to find. Likewise, repositories collect different data, but combining data from different datasets is difficult, and integrating data from other repositories is difficult also. If someone wants to analyze data, they would have to collect data from multiple sites, compile it themselves, and then run the data through another program for analysis. By hosting .csv files in our repository and analyzing the data, our program aims to remove the need for the user to compile data on their own.

2. Requirements

Requirement ID: uploadData

Requirement Type: functional

Requirement: Test software shall accept a .csv file and function must determine if file is written in the correct format (rows containing the times and columns containing the data, or vice versa) . Software will also be able to store a description of each file, its provider/owner, and what kind of access it has. (public/private)

Rationale: Program only reads, analyzes, and outputs .csv files. Program will also be able to tag documents with the meta-data.

Source: Requested by client

Acceptable Criterion: Software has a check to determine if .csv file is read and display error message if input is incorrect; must also be able to display meta-data

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: questionOne

Requirement Type: functional

Requirement: The software must be able to determine if there was a significant change in the target variable (answer Q1)

Rationale: Key feature of the program, and must answer questions provided by the client

Source: Requested by client

Acceptable Criterion: Execute software and ensure that elapsed time is no more than 1 second

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: questionTwo

Requirement Type: functional

Requirement: Program should be able to come up with the top-k variables that affect the target (answer Q2)

Rationale: Key feature of the program, and must answer questions provided by the client

Source: Requested by client

Acceptable Criterion: Execute software and ensure that elapsed time is no more than 1 second

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: questionThree

Requirement Type: functional

Requirement: Software must be able to forecast at the next timestamp with the target variable

Rationale: Key feature of the program and answers questions from the user

Source: Requested by client

Acceptable Criterion: Error margin is not greater than 50%

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: setTag

Requirement Type: functional

Requirement: Software will be able to set user-defined tags to documents

Rationale: Allows algorithm to quickly determine and search through repository for documents that correlate to target variable

Source: Requested by client

Acceptable Criterion: At least 3 tags on a document

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: setTarget

Requirement Type: functional

Requirement: Software will set a user-defined target variable as the target variable for the program

Rationale: Key feature to answer Q1, Q2, Q3

Source: Requested by client

Acceptable Criterion: User-defined variable is set as target variable

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: narrowData

Requirement Type: functional

Requirement: Software will narrow down the search of correlated variables and allow user to select which ones to include in search.

Rationale: Key feature to answer Q1, Q2, Q3

Source: Requested by client

Acceptable Criterion: Software will have no more than 10 correlating variables

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: visualizeData **Requirement Type:** functional

Requirement: Software will take in results from questionOne, questionTwo, and questionThree to display the results in graphical form.

Rationale: Key feature to answer Q1, Q2, Q3

Source: Requested by client

Acceptable Criterion: Software will present graphical results of Q1, Q2, Q3

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: metadata **Requirement Type:** functional

Requirement: Description of each variable, starting/ending timestamps, tag, and public/private setting.

Rationale: Allows selected pool of variables.

Source: Requested by client

Acceptable Criterion: The metadata will be stored in a database for easy and quick recover of data.

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: webApp **Requirement Type:** functional

Requirement: Website to host program

Rationale: Key feature of project

Source: Requested by client

Acceptable Criterion: A functional website that has a dedicated page for data providers and data customers to use, also has functional buttons.

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History:

Requirement ID: exportCSV

Requirement Type: functional

Requirement: Export data for future use

Rationale: The data computed by the program can be used on other higher level programs to answer more than just Questions 1, 2, 3.

Source: Requested by client

Acceptable Criterion: Users will be able to download the data in .csv files.

Dependencies: None

Conflicts: None

Supporting Materials:

Modification History: