COAL: Coal and Open-pit surface mining impacts on American Lands

Taylor Alexander Brown, Heidi Ann Clayton, and Xiaomei Wang

CS 461: Senior Capstone Fall 2016

Oregon State University

Abstract

Mining is known to cause environmental degradation, but software tools to identify mining impacts are lacking. Researchers studying this problem possess large imaging spectroscopy and environmental quality data sets as well as high-performance cloud-computing resources. This project provides a suite of algorithms using these data and resources to identify signatures of mining and correlate them with environmental impacts over time.

Contents

1 Introduction 3

1.1 Purpose 3

1.2 Scope 3

1.3 Definitions, acronyms, and abbreviations 3

1.4 References 3

1.5 Overview 3

2 Overall description 4

2.1 Product perspective 4

2.2 Product functions 4

2.3 User characteristics 4

2.4 Constraints 4

2.5 Assumptions and dependencies 4

3 Specific Requirements 4

3.1 Research background literature and libraries 4

3.2 Problem statement 4

3.3 Requirements document 5

3.4 Develop Software 5

3.4.1 Mineral identification and classification 5

3.4.2 Identify mining 5

3.4.3 Correlate mining with environmental impact 5

3.4.4 Rank and document changes over time 5

3.5 API documentation 5

3.6 Wiki 5

3.7 Homepage 5

3.8 Release Software 5

3.9 Publish Research Paper 5

1 INTRODUCTION

Our team name for this project is "COAL": Coal and Open-pit surface mining impacts on American Lands.

1.1 Purpose

The purpose of this document is to describe what COAL will provide as a piece of software. This document will touch on the interfaces, features, and constraints of COAL. This document is intended for developers, potential users of COAL, and any stakeholders in projects that will make use of or reference COAL.

1.2 Scope

COAL will be a suite of algorithms that identifies, classifies, and quantifies the effects of open-pit mining on the surrounding environment. This will be done by first using spectroscopy data to identify mining activity and then using data provided by GRaND to attempt to correlate these identified sites with environmental impacts over time. These algorithms will be provided as reusable components that can be used in cloud-based systems such as NASA Ames Stereo Pipeline. There will also be a website that will present the algorithms and related data visualization as well as links to the source code, project wiki, API documentation, and developer profiles to anyone interested in using COAL for their cloud-based platform or research.

1.3 Definitions, acronyms, and abbreviations

AGU	American Geophysical Union
API	Application Programming Interface
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
COAL	Coal and Open-pit surface mining impacts on American Lands
GRaND	Global Reservoir and Dam Database
Imaging Spectroscopy	"[A]cquisition of images where for each spatial resolution element in the image a spectrum of the energy arriving at the sensor is measured [1]."
LSE	Land and Solid Earth
MTM	Mountain-top Mining
NASA Ames Stereo Pipeline	A cloud-based platform that provides tools for processing stereo imagery from satellites, rovers, aerial photography, and historical images [2].

1.4 References

[1] "Imaging spectroscopy," in Jet Propulsion Laboratory, 2007. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.spectroscopy.html. Accessed: Nov. 4, 2016.

[2] "Neo-geography Toolkit," in National Aeronautics and Space Administration. [Online]. Available: https://ti.arc.nasa.gov/tech/asr/intelligent-robotics/ngt/stereo/. Accessed: Nov. 4, 2016.

1.5 Overview

The rest of this document contains an overall description of the project, a list of specific requirements, and signatures from the client and students.

2 Overall description

2.1 Product perspective

The COAL suite of algorithms will be compatible with cloud-based platforms. Users will interact with COAL's homepage in order to view a detailed analysis of COAL's algorithms and data visualization using spectroscopy and mineral maps. COAL's algorithms and data analysis will be interfaced with Python data analysis and visualization libraries. COAL's algorithms will have time and space complexity that is efficient enough for users to process very large data sets (e.g. 5+ GB). There are no hardware components to this product nor are there communications interfaces.

2.2 Product functions

Analyze imaging spectroscopy and environmental quality data sets to generate human-readable reports and visualizations. Provide API documentation, wiki, homepage, source code, and research to allow others to use, improve, and replicate this work.

2.3 User characteristics

The intended users are data scientists at research institutions. Users are expected to have access to image spectroscopy and environmental quality data sets which they wish to study. Users are also expected to have access to high-performance cloud computing resources and experience deploying applications to them. In general, the intended users are highly educated, experienced, and technical, but not necessarily proficient in programming.

2.4 Constraints

COAL will be efficient enough to process large data sets on high-performance cloud servers in a practical amount of time for research applications.

2.5 Assumptions and dependencies

The system requirements (such as operating system, build environment, and programming language interpreter version) are to be determined based on the requirements of the third-party libraries chosen to implement it.

3 Specific Requirements

3.1 Research background literature and libraries

On an ongoing basis for the duration of the project, review background literature and libraries and provide relevant references in project documentation. Literature may include videos, books, websites, and journal articles. Libraries include third-party software modules used and any new versions and functionalities that are added to them.

3.2 Problem statement

Provide documents that define the scope of the software suite from both the client perspective and the student perspective. The client problem statement provides a high-level overview of the client's goals for the project. The student problem statement describes the project from a computer science perspective, giving an overview of the problem domain and the steps taken to solve it.

3.3 Requirements document

Provide a document that describes in detail what tasks are to be completed and when.

3.4 Develop Software

Develop a suite of imagery processing algorithms using existing libraries and, if necessary, custom code. The behavior of each algorithm should be verified with automated tests. The algorithms should be efficient enough and effective enough for data scientists to justify adopting and deploying them to solve real-world problems. The results should include visualizations that allow data scientists to infer meaningful relationships from the data sets.

3.4.1 Mineral identification and classification

Implement algorithms to identify and classify minerals and their spatial relationships from imaging spectroscopy data.

3.4.2 Identify mining

Implement algorithms to identify regions of mining activity from the mineral identification and classification data.

3.4.3 Correlate mining with environmental impact

Implement algorithms to correlate the regions of mining activity with data on regions of environmental degradation.

3.4.4 Rank and document changes over time

Implement algorithms to derive spatial and temporal relationships between the regions of mining activity and the regions of environmental degradation.

3.5 API documentation

Provide API documentation for developers in a human-readable format that is automatically generated from the codebase.

3.6 Wiki

Provide a wiki with documentation for end users containing comprehensive usage instructions and relevant background information. Also provide a development blog with weekly updates from the development team.

3.7 Homepage

Provide a website that serves as the project homepage. The homepage should describe the project in general terms and demonstrate example imagery and data. Links to relevant sites including the source code repository, API documentation, Wiki, and developer profiles should be included. It should be accessible, interactive, and self-contained.

3.8 Release Software

Release the software as a collection of reusable modules which data scientists may deploy to cloud platforms. All code is to be released under the Apache License, Version 2.0.

3.9 Publish Research Paper

Publish a research paper and present it to a scientific society such as the AGU. The paper should explain the research problem, describe the software developed to solve it, and present as a case study an analysis of imagery from the AVIRIS project.

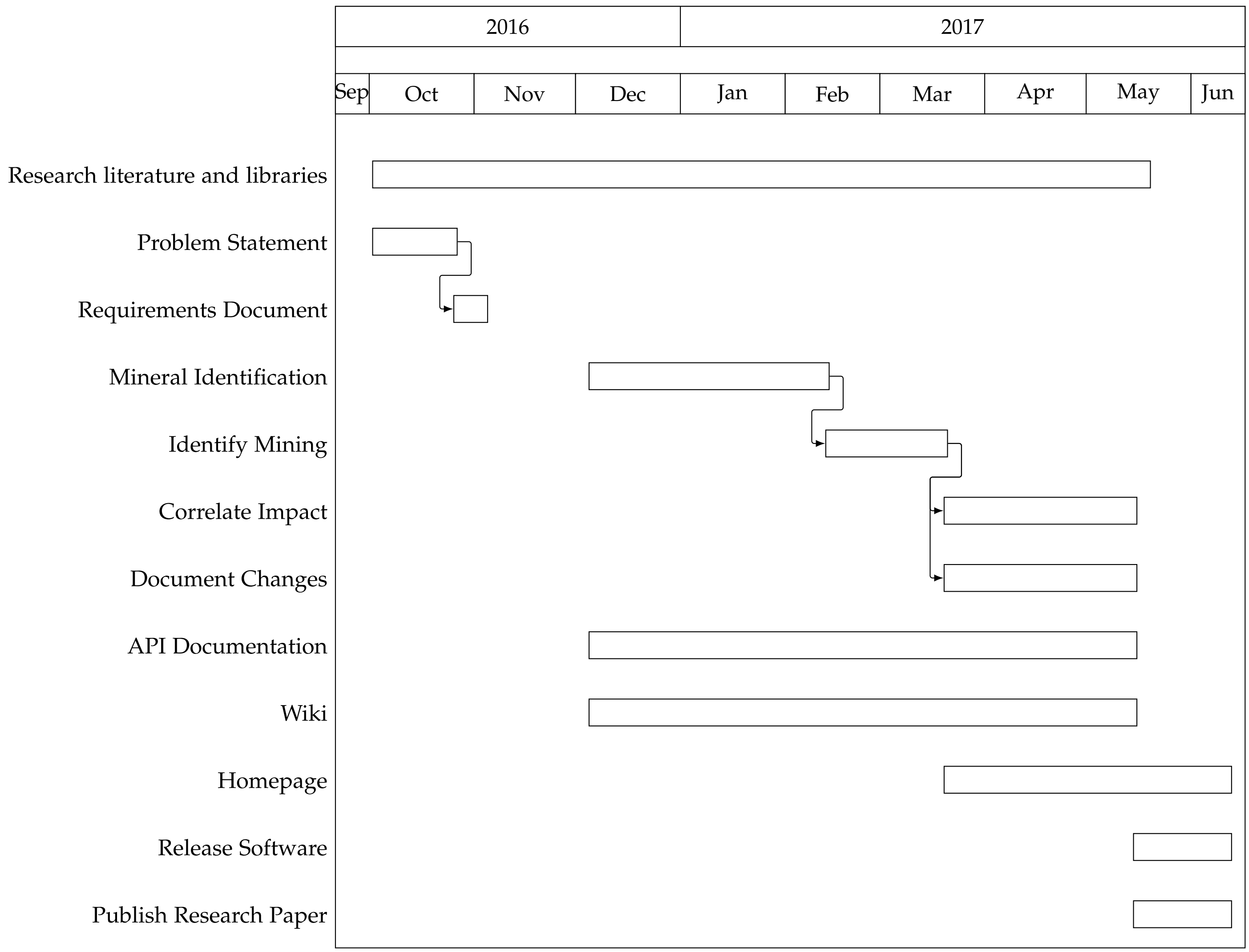

Fig. 1: Gantt Chart: Timeline of Project Tasks