Mail Crawler

Contents

[1. Introduction 3](#_Toc475074745)

[1.1. Unit Scope 3](#_Toc475074746)

[1.2. Related Checklist 3](#_Toc475074747)

[2. INTERFACES & ITEMs Used 4](#_Toc475074748)

[2.1. Interfaces – 4](#_Toc475074749)

[2.1.1. Jsoup-1.10.2.jar 4](#_Toc475074750)

[2.1.2. org.apache.commons.io.jar 4](#_Toc475074751)

[3. Solution Overview 5](#_Toc475074752)

[3.1. Assumptions 5](#_Toc475074753)

[3.2. High-Level Solution Description 5](#_Toc475074754)

[3.3. UML Diagram 5](#_Toc475074755)

[4. Code Changes 6](#_Toc475074756)

[4.1. Crawler 6](#_Toc475074757)

[4.2. IParser 6](#_Toc475074758)

[4.3. DocumentParser 6](#_Toc475074759)

[4.4. UriMapForYear 6](#_Toc475074760)

[4.5. Document 6](#_Toc475074761)

[4.6. IDownloader 6](#_Toc475074762)

[4.7. DocumentDownloader 6](#_Toc475074763)

[5. Testing Guidelines 7](#_Toc475074764)

[6. Open Issues 8](#_Toc475074765)

# Introduction

## Unit Scope

The purpose of this document is to describe the detailed design of the **Mail Crawler**. This document will cover the following topics:-

1. Requirement for Mail Crawler, and all its related information.
2. Assumptions.
3. List of Code changes.
4. Test scenario for the whole requirement.

## Related Checklist

N/A

# INTERFACES & ITEMs Used

## Interfaces –

### Jsoup-1.10.2.jar

### org.apache.commons.io.jar

# Solution Overview

## Assumptions

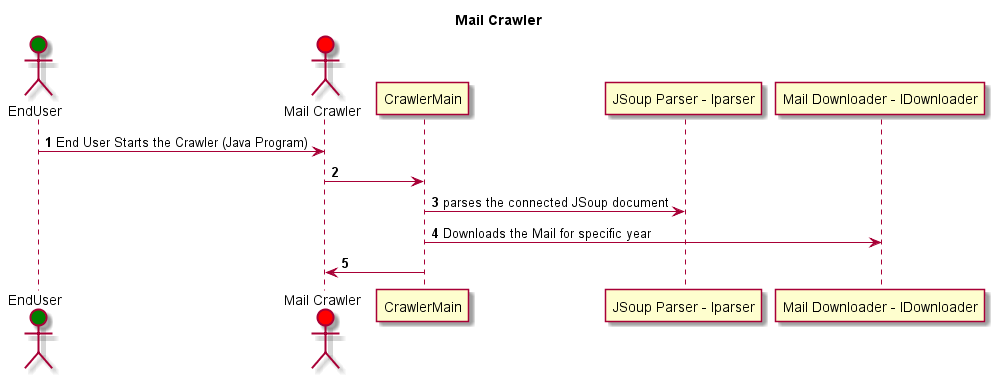
NA

## High-Level Solution Description

As part of this Development, following will be done:-

A simple crawler will be designed in Java to crawl and download all mails for a particular year (say 2015) from <http://mail-archives.apache.org/mod_mbox/maven-users/>

## UML Diagram



# Code Changes

## Crawler

Main class for initiating the process

## IParser

Abstract Interface for defining the parser methods

## DocumentParser

Concrete class for handling parsing mechanism of specified mail list

## UriMapForYear

Datatype that holds Map of all months with their relative URI’s

## Document

Class responsible for JSOUP library interaction for connecting to specified Mail list

## IDownloader

Abstract Interface for defining downloading methods

## DocumentDownloader

Concrete class for handling downloading mechanism of specified mail list

# Testing Guidelines

# Open Issues

Document Release Information

| Soft-ware Version | Editor | Edited Date | Comments |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| 1.0 | Navkaran Singh | March 19,2015 | Initial Version |