



This repository Search

Explore Gist Blog Help



terrywbrady

+ ▾



Georgetown-University-Libraries / File-Analyzer

forked from terrywbrady/File-Analyzer

Unwatch ▾

5

★ Unstar

13

Fork

7

File Analyzer Training Code4Lib 2015

Terry Brady edited this page 4 days ago · 11 revisions

Edit

New Page

Training Outline

- File Analyzer Overview
- Demonstration of Basic FileAnalyzer Tasks
- Customizing the FileAnalyzer for Every Department in the Library (Stories from Georgetown University Library)
- Demonstration -- Coding a File Analyzer Task
- Discussion -- What ideas do you have for the application?
- Try it yourself (optional)

Installation Notes (Optional)

The following preparation tasks will allow you to follow along with the training presentation. Since some session attendees may not have computers available or may not have the required software installed, this step will be optional.

1. Install and build the File Analyzer (required): [Installation instructions](#)
2. A Java IDE is recommended for last portion of the pre-conference. If you do not already have a Java IDE available, consider installing the Eclipse Standard Edition: <https://www.eclipse.org/downloads/>

File Analyzer Overview

- [File Analyzer Documentation](#)

Demonstration of Basic File Analyzer Tasks

User documentation is available at the link listed above.

- [User Interface - Search the File System](#)
- [User interface - viewing results](#)
- [Sorting results](#)
- [Filtering results](#)
- [Exporting results](#)
- [User interface - import records from a file](#)
- [User interface - Merging and Comparing Results](#)

Customizing the File Analyzer for

Pages 46

- Home
- File Analyzer Component Packages
- Installation instructions
- File Analyzer Stories
- File Analyzer Use Cases at Georgetown University
- Latest Features
- User Interface Overview
- Command Line Interface
- Batch Processing
- Coding new File Test Rules and new File Import Rules

- [File-Analyzer-Training-Code4Lib-2015](#)

Clone this wiki locally

<https://github.com/Georgetown>

Clone in Desktop

Every Department in the Library

- [File-Analyzer-Use-Cases-at-Georgetown-University](#)

Demonstration -- Coding a File Analyzer Task

Creating a File Test Rule or a File Import Rule

- [Coding new File Test Rules and new File Import Rules](#)

Coding a File Test Rule or File Import Rule

The project to be implemented will be determined by the interest of the group.

Parse MARC records and apply custom business logic

- [MARC-File-Analyzer](#)
- Sample MARC files: <https://github.com/code4lib/MARC-Records>

Analyze Digital Image Properties

- [Enhance the Image Properties Task](#)
- Some sample images: ([clone this wiki](#))

PDF Introspection

- [Enhance the Page Count Task](#)
- Some sample PDF's:

Discussion -- What ideas do you have for the application?

Try it yourself

Sample data files corresponding to these exercises will be provided at the start of the pre-conference session. Download the [exercise test files](#) from [GitHub](#). Extract the contents of the zip file after you download it.

Exercises to try

Run "Count Files by Type" on the "01_Flash Drive Inventory" folder.

- Sort the results from highest count to lowest count. What file type occurs most frequently?

Run "Match by Name" on the "01_Flash Drive Inventory" folder.

- Which file names have been duplicated?

- Remove your open tabs

Run "Match by Base Name"

- on the PDFs folder
- run it again on the Word Docs folder
- Which word document does not have a corresponding PDF?

Remove the tabs from all of your prior tests.

Run "Sort by Checksum" looking only at image files

- on the Checksum Tests folder.
- run it again on the Checksum Tests2 folder.
- Which files are not identical between the 2 folders?
- Remove the tab for your test on the Checksum Tests2 folder.
- Export the results from your first "Sort by Checksum" task as a tab-delimited file.
Export only the key and data fields.
- Import your checksum results using "Import Delimited File"
- Use the merge tool to compare your imported file to the results from your checksum test
- No differences should exist

+ Add a custom footer

