



This repository Search

[Explore](#) [Gist](#) [Blog](#) [Help](#)

terrywbrady

Georgetown-University-Libraries / **File-Analyzer**
forked from terrywbrady/File-Analyzer

Unwatch ▾

5

★ Unstar

13

Fork

7

Core File Import Rules

Terry Brady edited this page on Aug 18, 2014 · 7 revisions

Edit

New Page

[Core package contents](#)

Import Delimited File

This rule will import a delimited file (comma separated, tab separated, etc). Please specify the delimiter character to use.

Ⓜ Regular Expression Parser

This rule will parse each line of a file and add it to the results table. This rule requires an understanding of regular expressions.

File Import Action

Regular Expression Parser ▾

Importer Description Importer Properties Import Options

Regular Expression

Regular expression to test against each line of the file

^folder(....)(..)(...)*\$

Optional Output 1

Optional Output 1

Regex Field 1 ▾

Optional Output 2

Optional Output 2

Regex Field 3 ▾

Optional Output 3

Summary Counts				
Row	Pass/Fail	Data	Regex Field 1	Regex Field 3
1000000	FAIL	Key\Folder Name\Item FileThumb FileLicense Filedc....		Key\Folder Nam...
1000001	PASS	folder_00000eight.pdfLicense1.txta2012-11-01	_000	ei
1000002	PASS	folder_00001five.pdfab2012-11-01	_000	fi
1000003	PASS	folder_00002four.pdfcd2012-11-01	_000	fo
1000004	PASS	folder_00003one.pdfd2012	_000	on
1000005	PASS	folder_00004six.pdf2012-11-01	_000	si
1000006	PASS	folder_00005testfolderseven.pdf2012-11	_000	te
1000007	PASS	folder_00006three.pdfthree.pdf.jpgNo Date	_000	th
1000008	PASS	folder_00007two.pdf2012	_000	tw

Multi Parser

This rule will parse each line of a file against a sequence of regular expressions. Resulting values are stored in named matching groups. The sequence of patterns to apply to the file are stored in a rule file.

► Pages 46

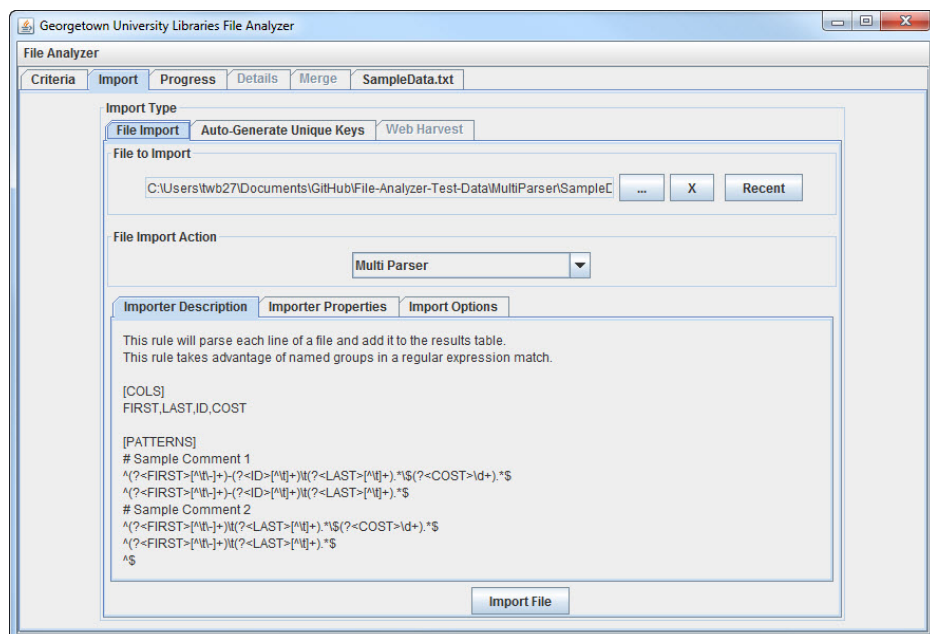
- Home
- File Analyzer Component Packages
- Installation instructions
- File Analyzer Stories
- File Analyzer Use Cases at Georgetown University
- Latest Features
- User Interface Overview
- Command Line Interface
- Batch Processing
- Coding new File Test Rules and new File Import Rules

- File-Analyzer-Training-Code4Lib-2015

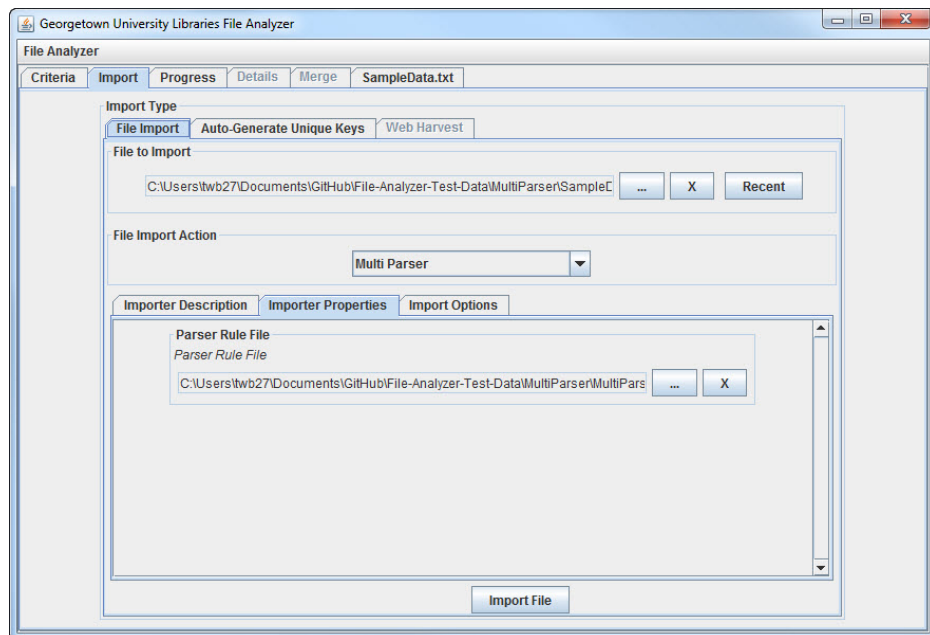
Clone this wiki locally

<https://github.com/Georgetown>

Clone in Desktop



On the properties tab, indicate the location of your Parser Rule File.



Here is a sample Parser Rule File

```
[COLS]
FIRST, LAST, ID, COST

[PATTERNS]
# Sample Comment 1
^(?<FIRST>[^\t-]+)-(?<ID>[^\t+])\t(?<LAST>[^\t+]).*\$(?<COST>[^\d+]).*$
^(?<FIRST>[^\t-]+)-(?<ID>[^\t+])\t(?<LAST>[^\t+]).*$
# Sample Comment 2
^(?<FIRST>[^\t-]+)\t(?<LAST>[^\t+]).*\$(?<COST>[^\d+]).*$
^(?<FIRST>[^\t-]+)\t(?<LAST>[^\t+]).*$
# Ignore empty lines
^$
```

If you run this rule against the following test data

```

Joe Smith   test $100 test
Jane   Doe $400 aa 22
Bob-F123   Foo
Dan-X222   Foober $671
this is a line with no tabs at all
Jim Davis  aaa $12345678

```

The resulting data will be displayed based on the named groups that were defined.

Row	Pass/Fail	Data	FIRST	LAST	ID	COST
1000000	PASS	JoeSmithtest \$100 test	Joe	Smith		100
1000001	PASS	JaneDoe\$400 aa 22	Jane	Doe		400
1000002	PASS	Bob-F123Foo	Bob	Foo	F123	
1000003	PASS	Dan-X222Foober\$671	Dan	Foobar	X222	671
1000004	FAIL	this is a line with no tabs at all				
1000005	PASS	Jim Davisaaa\$12345678	Jim Davis	aaa		12345678
1000006	PASS					
1000007	PASS					

Note: the sample files used in this example are available [here](#).

Count Key

Count the number of times a key value appears in a file.

File Import Action

Count Key

Importer Description | Importer Properties | Import Options

Delimiter
Delimiter character separating fields
Comma

HeadRow
Treat first row as header
Y

COL
Key Column starting at 1
16

Multiple? ▼

Summary Counts		
Key ▼	Count	Multiple?
China's search for ...	1	ONE
Children of God's fl...	1	ONE
Children's Rights ...	1	ONE
Child of the dark; t...	1	ONE
Chica da Silva : a ...	1	ONE
Chess fundament...	1	ONE
Chemistry: the cen...	4	MANY
Charting a new co...	1	ONE
Cassandra : a nov...	1	ONE
Case studies in bi...	2	MANY
Cases in compara...	1	ONE
Campus confidenti...	1	ONE
Caesar in Gaul an...	1	ONE
Byron and the rhet...	1	ONE
Byron and the politi	1	ONE

Counter Compliance - CSV

In the core package, the Counter Compliance tests apply only to text files. Use the updated version in the Demo package to parse XLSX files as well. The Counter Compliance Importer operates on a single file rather than on a collection of files.

See [Counter compliant reports](#)

+ Add a custom footer