

Georgetown-University-Libraries / File-Analyzer

forked from terrywbrady/File-Analyzer

Unwatch 5

★ Unstar 13

Fork 7

File Analyzer and Metadata Harvester <http://georgetown-university-libraries.github.io/File-Analyzer/> — Edit

221 commits

5 branches

8 releases

2 contributors

 branch: master



File-Analyzer / +










This branch is 56 commits ahead of terrywbrady:master

[Pull Request](#) [Compare](#)

Clean up combo box init warnings

 terrywbrady authored 6 days agolatest commit 769759667a 

 core	Clean up combo box init warnings	6 days ago
 demo	Handle worksheet iteration	2 months ago
 doc	move screen shots to ghpages	a year ago
 dspace	Remove dc.date.created test	7 days ago
 .gitignore	Merge pull request #1 from Georgetown-University-Libraries/master	10 months ago
 README	Merge pull request #40 from terrywbrady/master	10 months ago
 pom.xml	Marc Validator and Image Properties	a year ago

README

PURPOSE

Project Page: <http://georgetown-university-libraries.github.io/File-Analyzer/>

The File Analyzer and Metadata Harvester is a general purpose desktop (and command line) tool designed to automate simple, file-based operations. The File Analyzer assembles a toolkit of tasks a user can perform.

The tasks that have been written into the File Analyzer code base have been optimized for use by libraries, archives, and other cultural heritage institutions.

File Analyzer Wiki: <https://github.com/Georgetown-University-Libraries/File-Analyzer/wiki>

=====

This code has been derived from the NARA File Analyzer and Metadata Harvester which is available at <https://github.com/usnationalarchives/File-Analyzer>.

PREREQUISITES

- JDK 1.7 or higher (for build)
- JRE 1.7 or higher (for runtime)
- (If you need to run with Java 6, see Releases for an older version)
- Maven (or you will need to compile the modules manually)

INSTALLATION

- Clone this code to your computer
- Run `mvn install`
- Detailed Installation Instructions: <https://github.com/Georgetown-University-Libraries/File-Analyzer/wiki/Installation-instructions>

DEPLOYMENTS

This code will build 3 flavors of the File Analyzer.

(1) Core File Analyzer

Code

[Issues](#) 6[Pull Requests](#) 0[Wiki](#)[Pulse](#)[Graphs](#)[Settings](#)

HTTPS clone URL

<https://github.com/> 

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

[Clone in Desktop](#)[Download ZIP](#)

All code runs from a self-extracting jar file

(2) DSpace File Analyzer

This version of the file analyzer is a self-extracting jar file that references the core file analyzer jar file.

It contains tools for automating the creation of DSpace ingestion folders

(3) Demo File Analyzer

This version contains extensions illustrating various capabilities of the File Analyzer.

This version of the file analyzer is a self-extracting jar file that references both the core and dspace file analyzer jar files.

This version of the application uses features of Apache Tika and BagIt

=====
=====

License information is contained below.

Copyright (c) 2013, Georgetown University Libraries
All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

=====
=====

The NARA license for the base code is contained below.

NARA OPEN SOURCE AGREEMENT VERSION 1.3,
Based on NASA Open Source Agreement for Government Agencies, as approved by the Open Source Initiative.

THIS OPEN SOURCE AGREEMENT ("AGREEMENT") DEFINES THE RIGHTS OF USE, REPRODUCTION, DISTRIBUTION, MODIFICATION AND REDISTRIBUTION OF CERTAIN COMPUTER SOFTWARE ORIGINALLY RELEASED BY THE UNITED STATES GOVERNMENT AS REPRESENTED BY THE GOVERNMENT AGENCY LISTED BELOW ("GOVERNMENT AGENCY"). THE UNITED STATES GOVERNMENT, AS REPRESENTED BY GOVERNMENT AGENCY, IS AN INTENDED THIRD-PARTY BENEFICIARY OF ALL SUBSEQUENT DISTRIBUTIONS OR REDISTRIBUTIONS OF THE SUBJECT SOFTWARE. ANYONE WHO USES, REPRODUCES, DISTRIBUTES, MODIFIES OR REDISTRIBUTES THE SUBJECT SOFTWARE, AS DEFINED HEREIN, OR ANY PART THEREOF, IS, BY THAT ACTION, ACCEPTING IN FULL THE RESPONSIBILITIES AND OBLIGATIONS CONTAINED IN THIS AGREEMENT.

Government Agency: National Archives and Records Administration (NARA)

Government Agency Original Software Designation: org.nara.nwts.fileAnalyzer

Government Agency Original Software Title: NARA File Analyzer and Metadata Harvester.

User Registration Requested. Please Visit <http://www.archives.gov/social-media/github.html>.

Government Agency Point of Contact for Original Software: OpenGov@nara.gov.

1. DEFINITIONS

A. "Contributor" means Government Agency, as the developer of the Original Software, and any entity that makes a Modification.

B. "Covered Patents" mean patent claims licensable by a Contributor that are necessarily infringed by the use or sale of its Modification alone or when combined with the Subject Software.

C. "Display" means the showing of a copy of the Subject Software, either directly or by means of an image, or any other device.

D. "Distribution" means conveyance or transfer of the Subject Software, regardless of means, to another.

E. "Larger Work" means computer software that combines Subject Software, or portions thereof, with software separate from the Subject

Software that is not governed by the terms of this Agreement.

F. "Modification" means any alteration of, including addition to or deletion from, the substance or structure of either the Original Software or Subject Software, and includes derivative works, as that term is defined in the Copyright Statute, 17 USC 101. However, the act of including Subject Software as part of a Larger Work does not in and of itself constitute a Modification.

G. "Original Software" means the computer software first released under this Agreement by Government Agency with Government Agency designation National Archives and Records Administration (NARA) and entitled NARA File Analyzer and Metadata Harvester, including source code, object code and accompanying documentation, if any.

H. "Recipient" means anyone who acquires the Subject Software under this Agreement, including all Contributors.

I. "Redistribution" means Distribution of the Subject Software after a Modification has been made.

J. "Reproduction" means the making of a counterpart, image or copy of the Subject Software.

K. "Sale" means the exchange of the Subject Software for money or equivalent value.

L. "Subject Software" means the Original Software, Modifications, or any respective parts thereof.

M. "Use" means the application or employment of the Subject Software for any purpose.

2. GRANT OF RIGHTS

A. Under Non-Patent Rights: Subject to the terms and conditions of this Agreement, each Contributor, with respect to its own contribution to the Subject Software, hereby grants to each Recipient a non-exclusive, world-wide, royalty-free license to engage in the following activities pertaining to the Subject Software:

1. Use
2. Distribution
3. Reproduction
4. Modification
5. Redistribution
6. Display

B. Under Patent Rights: Subject to the terms and conditions of this Agreement, each Contributor, with respect to its own contribution to the Subject Software, hereby grants to each Recipient under Covered Patents a non-exclusive, world-wide, royalty-free license to engage in the following activities pertaining to the Subject Software:

1. Use
2. Distribution
3. Reproduction
4. Sale
5. Offer for Sale

C. The rights granted under Paragraph B. also apply to the combination of a Contributor's Modification and the Subject Software if, at the time the Modification is added by the Contributor, the addition of such Modification causes the combination to be covered by the Covered Patents. It does not apply to any other combinations that include a Modification.

D. The rights granted in Paragraphs A. and B. allow the Recipient to sublicense those same rights. Such sublicense must be under the same terms and conditions of this Agreement.

3. OBLIGATIONS OF RECIPIENT

A. Distribution or Redistribution of the Subject Software must be made under this Agreement except for additions covered under paragraph 3H.

1. Whenever a Recipient distributes or redistributes the Subject Software, a copy of this Agreement must be included with each copy of the Subject Software; and

2. If Recipient distributes or redistributes the Subject Software in any form other than source code, Recipient must also make the source code freely available, and must provide with each copy of the Subject Software information on how to obtain the source code in a reasonable manner on or through a medium customarily used for software exchange.

B. Each Recipient must ensure that the following copyright notice appears prominently in the Subject Software:

Copyright 2011

United States Government as represented by Archivist of the United States, National Archives and Records Administration. No copyright is claimed in the United States under Title 17, U.S.Code. All Other Rights Reserved.

C. Each Contributor must characterize its alteration of the Subject Software as a Modification and must identify itself as the originator

of its Modification in a manner that reasonably allows subsequent Recipients to identify the originator of the Modification. In fulfillment of these requirements, Contributor must include a file (e.g., a change log file) that describes the alterations made and the date of the alterations, identifies Contributor as originator of the alterations, and consents to characterization of the alterations as a Modification, for example, by including a statement that the Modification is derived, directly or indirectly, from Original Software provided by Government Agency. Once consent is granted, it may not thereafter be revoked.

D. A Contributor may add its own copyright notice to the Subject Software. Once a copyright notice has been added to the Subject Software, a Recipient may not remove it without the express permission of the Contributor who added the notice.

E. A Recipient may not make any representation in the Subject Software or in any promotional, advertising or other material that may be construed as an endorsement by Government Agency or by any prior Recipient of any product or service provided by Recipient, or that may seek to obtain commercial advantage by the fact of Government Agency's or a prior Recipient's participation in this Agreement.

F. In an effort to track usage and maintain accurate records of the Subject Software, each Recipient, upon receipt of the Subject Software, is requested to provide Government Agency, by e-mail to the Government Agency Point of Contact listed in clause 5.F., the following information: User name, Title, Organization name, E-mail address. Recipient's name and personal information shall be used for statistical purposes only. Once a Recipient makes a Modification available, it is requested that the Recipient inform Government Agency, by e-mail to the Government Agency Point of Contact listed in clause 5.F., how to access the Modification.

G. Each Contributor represents that that its Modification is believed to be Contributor's original creation and does not violate any existing agreements, regulations, statutes or rules, and further that Contributor has sufficient rights to grant the rights conveyed by this Agreement.

H. A Recipient may choose to offer, and to charge a fee for, warranty, support, indemnity and/or liability obligations to one or more other Recipients of the Subject Software. A Recipient may do so, however, only on its own behalf and not on behalf of Government Agency or any other Recipient. Such a Recipient must make it absolutely clear that any such warranty, support, indemnity and/or liability obligation is offered by that Recipient alone. Further, such Recipient agrees to indemnify Government Agency and every other Recipient for any liability incurred by them as a result of warranty, support, indemnity and/or liability offered by such Recipient.

I. A Recipient may create a Larger Work by combining Subject Software with separate software not governed by the terms of this agreement and distribute the Larger Work as a single product. In such case, the Recipient must make sure Subject Software, or portions thereof, included in the Larger Work is subject to this Agreement.

J. Notwithstanding any provisions contained herein, Recipient is hereby put on notice that export of any goods or technical data from the United States may require some form of export license from the U.S. Government. Failure to obtain necessary export licenses may result in criminal liability under U.S. laws. Government Agency neither represents that a license shall not be required nor that, if required, it shall be issued. Nothing granted herein provides any such export license.

4. DISCLAIMER OF WARRANTIES AND LIABILITIES; WAIVER AND INDEMNIFICATION

A. No Warranty: THE SUBJECT SOFTWARE IS PROVIDED "AS IS" WITHOUT ANY WARRANTY OF ANY KIND, EITHER EXPRESSED, IMPLIED, OR STATUTORY, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTY THAT THE SUBJECT SOFTWARE WILL CONFORM TO SPECIFICATIONS, ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR FREEDOM FROM INFRINGEMENT, ANY WARRANTY THAT THE SUBJECT SOFTWARE WILL BE ERROR FREE, OR ANY WARRANTY THAT DOCUMENTATION, IF PROVIDED, WILL CONFORM TO THE SUBJECT SOFTWARE. THIS AGREEMENT DOES NOT, IN ANY MANNER, CONSTITUTE AN ENDORSEMENT BY GOVERNMENT AGENCY OR ANY PRIOR RECIPIENT OF ANY RESULTS, RESULTING DESIGNS, HARDWARE, SOFTWARE PRODUCTS OR ANY OTHER APPLICATIONS RESULTING FROM USE OF THE SUBJECT SOFTWARE. FURTHER, GOVERNMENT AGENCY DISCLAIMS ALL WARRANTIES AND LIABILITIES REGARDING THIRD-PARTY SOFTWARE, IF PRESENT IN THE ORIGINAL SOFTWARE, AND DISTRIBUTES IT "AS IS."

B. Waiver and Indemnity: RECIPIENT AGREES TO WAIVE ANY AND ALL CLAIMS AGAINST THE UNITED STATES GOVERNMENT, ITS CONTRACTORS AND

SUBCONTRACTORS, AS WELL AS ANY PRIOR RECIPIENT. IF RECIPIENT'S USE OF THE SUBJECT SOFTWARE RESULTS IN ANY LIABILITIES, DEMANDS, DAMAGES, EXPENSES OR LOSSES ARISING FROM SUCH USE, INCLUDING ANY DAMAGES FROM PRODUCTS BASED ON, OR RESULTING FROM, RECIPIENT'S USE OF THE SUBJECT SOFTWARE, RECIPIENT SHALL INDEMNIFY AND HOLD HARMLESS THE UNITED STATES GOVERNMENT, ITS CONTRACTORS AND SUBCONTRACTORS, AS WELL AS ANY PRIOR RECIPIENT, TO THE EXTENT PERMITTED BY LAW. RECIPIENT'S SOLE REMEDY FOR ANY SUCH MATTER SHALL BE THE IMMEDIATE, UNILATERAL TERMINATION OF THIS AGREEMENT.

5. GENERAL TERMS

A. Termination: This Agreement and the rights granted hereunder will terminate automatically if a Recipient fails to comply with these terms and conditions, and fails to cure such noncompliance within thirty (30) days of becoming aware of such noncompliance. Upon termination, a Recipient agrees to immediately cease use and distribution of the Subject Software. All sublicenses to the Subject Software properly granted by the breaching Recipient shall survive any such termination of this Agreement.

B. Severability: If any provision of this Agreement is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Agreement.

C. Applicable Law: This Agreement shall be subject to United States federal law only for all purposes, including, but not limited to, determining the validity of this Agreement, the meaning of its provisions and the rights, obligations and remedies of the parties.

D. Entire Understanding: This Agreement constitutes the entire understanding and agreement of the parties relating to release of the Subject Software and may not be superseded, modified or amended except by further written agreement duly executed by the parties.

E. Binding Authority: By accepting and using the Subject Software under this Agreement, a Recipient affirms its authority to bind the Recipient to all terms and conditions of this Agreement and that that Recipient hereby agrees to all terms and conditions herein.

F. Point of Contact: Any Recipient contact with Government Agency is to be directed to the designated representative as follows:

_____.

