

JWAT v1.0.0 (2013-02-11)

Released 2013-02-11.

- Highlights
 - ARC package
 - The ArcReader was rewritten leaving to code unturned
 - Unit testing of the remaining ARC classes
 - ArcWriter usable
 - ArcFile detection improved with an ArcRecord detection
 - WARC package
 - WarcProfile now treated as an URI and not a string
 - Common package
 - Custom URI implementation added including profiles
 - A custom URI implementation was added due to the fact that many URIs used on the web are not valid according to the RFC. UriProfiles were added to allow for different levels of URIs leniency.
 - Full list of issues resolved in this release.

Highlights

ARC package

The ArcReader was rewritten leaving to code unturned

Version header in a Version block record now treated as a payload header, exactly like a HTTP header for ARC records.

Much of the ARC code was also restructured for this release.

Unit testing of the remaining ARC classes

Added unit tests so all classes are now covered.

ArcWriter usable

The ARC writer is now working.

ArcFile detection improved with an ArcRecord detection

Only checking for "filedesc:" in the beginning of an ARC file was not lenient enough, so methods for checking for a valid scheme at the beginning of a stream was added.

WARC package

WarcProfile now treated as an URI and not a string

The WARC code was reviewed and minor improvements where made.

Common package

Custom URI implementation added including profiles

A custom URI implementation was added due to the fact that many URIs used on the web are not valid according to the RFC. UriProfiles were added to allow for different levels of URIs leniency.

Full list of issues resolved in this release.

T	Key	P	Summary
	JWAT-56		Wget WARC files declared as Invalid
	JWAT-8		Unit tests and coverage of ArcRecordBase, ArcRecord and ArcVersionBlock
	JWAT-64		ArcReaderFactory.isArcFile only check for filedesc:
	JWAT-61		Warc-Profile is validated as a string not an URI

JWAT-60	Rewrite of URI validation in ARC header does not handle no-type before URI check
JWAT-57	Unit test in jwt-common fails occasionally when it should pass, due to ordering of results.
JWAT-58	The recoded ARC validation is too strict for IA ARCs
JWAT-59	Improved URI validation is too strict
JWAT-45	ARC writer
JWAT-46	ARC reader refactoring
JWAT-38	NullPointerException if the versionblock is only one line.
JWAT-35	ARC reader should probably keep the raw values instead of the converted/validated

12 issues

Downloads



Artifacts

