

JWAT v0.8.0 (2012-02-09)

Released 2012-02-09.

- [Highlights](#)
 - [GZip reader/validator](#)
 - [ARC reader/validator](#)
 - [Read and validate v1.0 and v2.0 ARC files](#)
 - [v1.1 extension supported.](#)
 - [Payload wrapper](#)
 - [HttpResponse supported](#)
 - [WARC reader/validator](#)
 - [Multi-line headers are supported](#)
 - [All WARC headers are parsed and validated](#)
 - [WARC headers present in a record are validated according to the record's Warc-Type](#)
 - [Payload wrapper](#)
 - [HttpResponse supported](#)
- [Full list of issues resolved in this release.](#)

Highlights

GZip reader/validator

Gives the ability to read a GZip file comprising of one or more concatenated GZip entries. Each entry's header is validated against the RFC specifications and the entries compressed data is made available through an `InputStream`. Will be refactored for v1.0.

ARC reader/validator

[Read and validate v1.0 and v2.0 ARC files](#)

v1.0 and v2.0 files are parsed and validated according to the specification.

[v1.1 extension supported.](#)

The semi-official v1.1 format is also supported.

[Payload wrapper](#)

Payload is supported and exposed to the end-user by means of a wrapper.

[HttpResponse supported](#)

HttpResponses are detected and parsed/validated and exposed to the end-user.

WARC reader/validator

[Reads and validates WARC version 1.0 files. \(Not all parts of the ISO are covered yet\)](#)

[Multi-line headers are supported](#)

Multi-line headers are parsed and validated with the exception of UTF-8, quoted-strings and encoded words.

[All WARC headers are parsed and validated](#)

All WARC headers are parsed and validated according to type. (String, Numeric, URI, Date, IP, Content-Type, Digest)

[WARC headers present in a record are validated according to the record's Warc-Type](#)

Each WARC header is validated according to the WARC-Type of the record it is located in. According to the ISO standard header presence can either be: must, must not, shall, shall not, may or ignore.



[Payload wrapper](#)

Payload is supported and exposed to the end-user by means of a wrapper.

[HttpResponse supported](#)

HttpResponses are detected and parsed/validated and exposed to the end-user.

Full list of issues resolved in this release.

T	Key	P	Summary
	JWAT-15		Basic WARC reader/validator implementation conforming to most of the ISO standard

	JWAT-21		Refactor HttpResponse for use in ARC and WARC. Add support for Digesting.
	JWAT-18		Digest validation in WARC reader
	JWAT-16		Digest support in ARC reader
	JWAT-14		Unit tests and coverage of most of the WARC code (except WarcRecord and WarcWriter related classes)
	JWAT-7		Base16, 32 and 64 added for use in Digesting. Base2 added because. Unit tests included.
	JWAT-6		Refactor Payload for use in ARC and WARC. Add support for Digesting.
	JWAT-4		Move ARC Payload, HttpResponse and stream classes to subpackage
	JWAT-5		Streams refactored and a few news ones added
	JWAT-2		Unit tests and coverage of most of the ARC code (except ArcRecordBase, ArcRecord and ArcVersionBlock)
	JWAT-3		GZip code extracted from BnF module and placed in subpackage
	JWAT-1		Basic ARC reader/validator refactored from BnF code
	JWAT-37		Expose tri-state boolean with digest comparison status for a warc record
	JWAT-34		Expose Computed Digest Values in ARC and WARC readers
	JWAT-32		Save and expose HttpResponse headers
	JWAT-30		Maven - Cleanup pom and prepare for 1st sonatype release
	JWAT-25		Unit tests and coverage of readLine methods across all stream implementations
	JWAT-24		Unit tests and coverage of RandomFileAccessl/OStreams in common
	JWAT-26		Unit tests and coverage of Payload and HttpResponse
	JWAT-23		Primitive WARC writer

Showing 20 out of 22 issues

Downloads

Artifacts