JWAT v0.9.0 (2012-05-11)

Released 2012-05-11.

Highlights

GZip package

Refactored GZip reader/validator

The reader/validator was completely rewritten since the original one was badly implemented.

GZip writer

A GZip writer was implemented for completeness and also to test the reader fully.

Input/Output stream

The reader/writer now has support for streams directly.

Validation reporting

Warnings/Errors are now reported in the same manor as the ARC/WARC packages.

Unit tests completeness

Added more unit tests so also all GZip code is test covered.

ARC/WARC

More unit tests

Added better test coverage but still not completely covered.

Unified the reader API

The ARC/WARC factories/readers should now have a similar API.

Common package

Unit testing of Stream classes

The stream classes were not completely unit tested in the previous release, but they should be now.

Completed the HeaderLineReader

The HeaderLineReader now includes validation of RAW, US-ASCII, ISO-8859-1, UTF-8, Quoted-Strings and Encoded Words.

Full list of issues resolved in this release.

T	Key	P	Summary
	JWAT-20		Unit tests and coverage for utf-8, quoted-strings and encoded words support.
	JWAT-19		Complete WARC headerline reader support for utf-8, quoted-strings and encoded words.
	JWAT-13		Unit tests and coverage of GZip code
	JWAT-11		Refactor HttpResponse parser code
	JWAT-9		Refactor GZip code
	JWAT-42		Refactor and unify gzip, arc and warc error/warning handling
	JWAT-41		ArcReader/Record does not seems to return errors/warnings
	JWAT-40		GZip report errors/warnings
	JWAT-39		HttpResponse uses mark/reset which have been disabled in the refactored streams.
	JWAT-29		Unit tests and coverage of streams in streams (validate consumption)
	JWAT-27		Unit tests and coverage of HttpResponse in a Payload
	JWAT-22		Unify ARC and WARC Reader/Factory Api
	JWAT-10		Implement GZip writer

13 issues

Downloads

Artifacts