

Documentation

The pages below describe the individual packages and also the process by which ARC and WARC files are read and validated.

If you can not find the information you seek you can always try the javadocs or look at the source code. As a last resort you are also welcome to email me.

Package layout

This toolkit includes the following packages:

- `jwt-common`: General purpose classes including specialized streams, binary->string encoding and common arc/warc http-response/payload code.
- `jwt-gzip`: GZip reader/validator/writer, including input/output streams for data.
- `jwt-arc`: Contains Arc reader/validator/writer specific classes.
- `jwt-warc`: Contains Warc reader/validator/writer specific classes.

`jwt-common`

common package

`jwt-gzip`

gzip package

`jwt-arc`

arc package

`jwt-warc`

warc package

ARC reader process

Describes the steps taken to read and validate an ARC record.

WARC reader process

Describes the steps taken to read and validate a WARC record.

Caveat user

Known side-effects and pitfalls from the current reading/validating strategy