

jwat-warc

warc package

org.jwat.warc:

WarcConstants.java

Most of the constants should be collected in this class, most of which are primarily for internal use.

ReaderFactory and Readers

- **WarcReaderFactory.java:** This factory can be used to create the various types of readers with optional buffering. You can either get **compressed** or **uncompressed** readers. There are also methods which can **auto-detect** whether or not a compressed reader is required.
- **WarcReader.java:** Abstract reader class which is the base for the all the readers. It also defines the options which can be set on a reader. Currently only digest options.
- **WarcReaderCompressed.java:** A reader implementation for reading compressed records.
- **WarcReaderUncompressed.java:** A reader implementation for reading uncompressed records.

WarcRecord.java

This class contains the record parser, fields and validation.

Auxiliary classes

- **WarcHeaderLine.java:** Reading a WARC header encapsulates each line in instances of this class.
- **WarcDateParser.java:** Parses and validated an WARC date.
- **WarcDigest.java:** Parses, validates and encapsulates a WARC digest header (algorithm, digest, encoding). The encoding is auto-detected and added later in the reading process.
- **WarcErrorType.java:** Defines the different possible error types.
- **WarcValidationError.java:** Defines an WARC validation error using a type, key and value.

Writers

- **WarcWriter.java:** Abstract writer class which is the base for all the writers.
- **WarcWriterUncompressed.java:** A writer implementation prototype.