

Usage

Since this project is mainly aimed at building a general purpose Web Archiving Toolkit these packages do not contain any applications, instead they are intended to be used as building blocks.

Used as a Maven Dependency

If you want to use JWAT from any build environment that supports artifacts, the following can be used.

pom.xml

```
<dependencies>
  <dependency>
    <groupId>org.jwat</groupId>
    <artifactId>jwat-common</artifactId>
    <version>1.0.4</version>
  </dependency>

  <dependency>
    <groupId>org.jwat</groupId>
    <artifactId>jwat-gzip</artifactId>
    <version>1.0.4</version>
  </dependency>

  <dependency>
    <groupId>org.jwat</groupId>
    <artifactId>jwat-arc</artifactId>
    <version>1.0.4</version>
  </dependency>

  <dependency>
    <groupId>org.jwat</groupId>
    <artifactId>jwat-warc</artifactId>
    <version>1.0.4</version>
  </dependency>

  <dependency>
    <groupId>org.jwat</groupId>
    <artifactId>jwat-archive-common</artifactId>
    <version>1.0.4</version>
  </dependency>
</dependencies>
```

Used as jars.

If you want to use the jars by themselves you can just download them and place them in the classpath.

You should be able to find them on the [Sonatype Maven Repository](https://oss.sonatype.org/content/repositories/releases/org/jwat/). Search for "jwat".

Released artifacts are available directly from here: <https://oss.sonatype.org/content/repositories/releases/org/jwat/>

Snapshot artifacts are available directly from here: <https://oss.sonatype.org/content/repositories/snapshots/org/jwat/>

Sequential or Random Access usage:

The ARC/WARC/GZip readers can be used to read either all the records/entries in a file sequentially or select records/entries in random order.

Both scenarios are supported by the various factory and reader methods.

Compression:

The ARC/WARC factory and reader methods support both compressed and uncompressed files in varying combinations.

Basically you can create a reader for compressed or uncompressed files, but the factory also includes an auto-detection mechanism which can determine the correct reader on a per file basis.

Writing compressed ARC/WARC files is also possible though the use of different methods in the writer factories.

GZip compression is only supported on ARC/WARC files where each record is compressed individually and concatenated into one file and not the case where the whole ARC/WARC file and all it's records are GZip'ed as a whole. The later mostly because this makes random access to individual record highly ineffective.

Examples

Code examples

- [Reading a WARC file](#) — Insert short page description here