

ARC reader process

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

Both parsers work given the same input. Normally a version block is read first and after that a varying number of records are read until End-Of-File is encountered.

Steps to parsing an ARC version block:

The following steps are taken when parsing an ARC version block:

1. The reader starts by reading 3 lines. (**recordLine**, **versionLine** and **fieldLine**)
2. If the **recordLine** is non empty, it check for a leading "filedesc://".
3. If the **versionLine** is non empty, parse and validate the version, reserved and origin fields.
4. If the **fieldLine** is non empty, identify which field version is being used. (Defaults to v1.0 is unidentified)
5. If the **recordLine** is non empty, parse and validate all the record fields.
6. Process payload, if present.

Step 1

The first thing the version block reader does is initialize various internal fields and then read 3 lines from the input stream. The 3 lines read are (in order): **recordLine**, **versionLine** and **fieldLine**.

Step 2

Next thing the reader does is to verify that the **recordLine** starts with "filedesc://" which would identify the beginning of a version block. If this is not the case an error is logged.

Step 3

The reader checks for a non empty **versionLine** and proceeds by splitting it up and validating the individual fields.

Step 4

The **fieldLine** is fixed for each version. So we compare the line and decide which field scheme to use on the following records.

Step 5

Since the reader now has the version and field information its time to parse the **recordLine** itself. The **recordLine** is split into individual fields which are parse and validated according to the field definition (**fieldLine**).

After the header has been processed errors are checked and the compliance status is updated.

Step 6

Lastly the reader checks the record length and reports if it is missing or too small. If there is still record data left the remaining data is sent to the payload processor. Any data present is saved in a string value. (The **ARC v1.1** specification added an xml payload to the version block)

Steps to parsing an ARC record:

The following steps are taken when parsing an ARC record:

1. Read a single line (**recordLine**), if it is empty read the next one until it is non empty.
2. Check for a non empty **recordLine** and if so parse and validate the fields according to the ARC file version.
3. Process payload, if present.

Step 1

Initialize various internal fields and repeatedly read the next line on the stream (**recordLine**) until it is non empty or the End-Of-File is

encountered.

Step 2

If the **recordLine** was non empty the line is split into individual fields which are parsed and validated according to the field definition in the version block.

After the header has been processed errors are checked and the compliance status is updated.

Step 3

Last step is to check the record length and report if it is missing or too small. If there is still record data left the remaining data is sent to the payload processor. The payload processor wraps the content in an object which is exposed to the end user.

If the payload processor detects a **http response** header this is also parsed, validated and exposed to the end user.