



Metadata Extraction Tool Changes

Version: 3.0.

Metadata Extraction Tool

Changes

Version: 3.0.

Table of Contents

✓	Introduction	3
✓	Version 3.0	3
	File Support.....	3
	Major Changes	3

✓ **Introduction**

This document covers the changes to the Metadata Extraction Tool.

✓ **Version 3.0**

Version 3.0 is the initial release to the Open Source community.

File Support

Version 3.0 of the Metadata Extraction Tool changes the way the files are identified by the tool. In earlier versions, the Metadata Extractor tool relied heavily on extensions. The new version relies on magic numbers where possible. Notable exceptions to this rule are:

- MP3 files are still detected by their extension as using the magic number (0xFF at the start of the file) risks a large number of false positives.
- Microsoft Office and Works documents are identified based on their extension, plus a check to ensure that they are valid OLE files. This is due to difficulties in being able to consistently locate metadata that identifies the type of the document. While this information is available, it varies in its position.
- There have been major changes to the XMLAdapter. The new version will extract the information from the XML and DOCTYPE declarations, while previous versions essentially copied the source XML file.

Major Changes

Version 3.0 includes the following major changes:

- Additional JavaDocs, especially in the interfaces where development is expected.
- New nz.govt.natlib.samples package containing usage samples.
- Update and addition of new documentation.
- Changes to the XMLAdapter to extract the XML version and encoding types and extract information from the DOCTYPE header.
- Creation of ANT build scripts to automate the build process.
- Modification of the default build to include all adapters.