

Deliverable 1:

Identify relevant data sources from where a file type information (as described above) can be extracted based on filename or file extension. List at least 5 relevant sources and explain the rationale on why it should be used.

1. File-extensions

2. FileInfo

3. Apache Tika

4. FILExt

5. DotWhat

6. OpenWith

File-Extensions.org

It is a library containing thousands of file extensions and the database is still growing. A large number of file type entries have detailed descriptions, including their current use and the list of programs that can open, view, edit, convert or play unknown file you search for.

Link: <https://www.file-extensions.org>

FileInfo.com

It contains a searchable database of over 10,000 file extensions with detailed information about the associated file types. We can look up information about unknown file types and find programs that open the files. It has become the authoritative website where developers can submit new file extensions and provide information about file types.

Link: <https://fileinfo.com/>

Apache Tika

The Apache Tika™ toolkit detects and extracts metadata and text from over a thousand different file types (such as PPT, XLS, and PDF). All of these file types can be parsed through a single interface, making Tika useful for search engine indexing, content analysis, translation, and much more. Apache Tika is a content detection and analysis framework that is written in Java and stewarded at the Apache Software Foundation.

Link: <https://tika.apache.org/>

FILExt.com

FILExt is one of the oldest and most respected collections of file formats and file extensions. It helps us to find various extensions and information related to it. It also helps to identify, open, view, or convert unknown files.

Link: <https://filext.com/>

DotWhat.net

It is a database of file extension information, covering multiple operating systems from Microsoft's Windows, Apple's OS X and all variations of UNIX to those used on the latest mobile devices and phones. It provide various information on each file extension. Also reviews and add new extension provided by site visitors.

Link: <http://dotwhat.net/>

OpenWith.org

OpenWith provides detailed information about most file extension and links to free programs that can open and create each type of file. It provides data about various type of extension present in its database. We can look up information about unknown file types and find programs that open the files.

Link: <https://www.openwith.org/>