

TIKA - FILE FORMATS

http://www.tutorialspoint.com/tika/tika_file_formats.htm

Copyright © tutorialspoint.com

File Formats Supported by Tika

The following table shows the file formats Tika supports.

File format	Package Library	Class in Tika
XML	org.apache.tika.parser.xml	XMLParser
HTML	org.apache.tika.parser.html and it uses Tagsoup Library	HtmlParser
MS-Office compound document Ole2 till 2007 ooxml 2007 onwards	org.apache.tika.parser.microsoft org.apache.tika.parser.microsoft.ooxml and it uses Apache Poi library	OfficeParserole2 OOXMLParserooxml
OpenDocument Format openoffice	org.apache.tika.parser.odf	OpenOfficeParser
portable Document FormatPDF	org.apache.tika.parser.pdf and this package uses Apache PdfBox library	PDFParser
Electronic Publication Format digitalbooks	org.apache.tika.parser.epub	EpubParser
Rich Text format	org.apache.tika.parser.rtf	RTFParser
Compression and packaging formats	org.apache.tika.parser.pkg and this package uses Common compress library	PackageParser and CompressorParser and its sub-classes
Text format	org.apache.tika.parser.txt	TXTParser
Feed and syndication formats	org.apache.tika.parser.feed	FeedParser
Audio formats	org.apache.tika.parser.audio and org.apache.tika.parser.mp3	AudioParser MidiParser Mp3- for mp3parser
Imageparsers	org.apache.tika.parser.jpeg	JpegParser-for jpeg images
Videoformats	org.apache.tika.parser.mp4 and org.apache.tika.parser.video this parser internally uses Simple Algorithm to parse flash video formats	Mp4parser FlvParser
java class files and jar files	org.apache.tika.parser.asm	ClassParser CompressorParser
Mobxformat emailmessages	org.apache.tika.parser.mbox	MobXParser
Cad formats	org.apache.tika.parser.dwg	DWGPParser
FontFormats	org.apache.tika.parser.font	TrueTypeParser
executable programs	org.apache.tika.parser.executable	ExecutableParser

and libraries

Loading [MathJax]/jax/output/HTML-CSS/fonts/TeX/fontdata.js