

TikaParseContext

Using [ParseContext](#) to Control Parsing

The ParseContext is used to configure parsing for a given file.

The general use is

```
parseContext.set(MyClass.class, new MyClass());  
parser.parse(inputStream, contentHandler, metadata, parseContext);
```

General

The following uses apply to several parsers:

1. Handling embedded files

- 1a. EmbeddedDocumentExtractor – for handling embedded files, the user can specify a custom EmbeddedDocumentExtractor.

- 1b. Parser – if the user fails to pass in an EmbeddedDocumentExtractor, the parsers will look for a Parser.class in the [ParseContext](#), and Tika will build a ParsingEmbeddedDocumentExtractor based on that Parser automatically.

- 1c. NOTE: As of Tika 1.15, if the user doesn't specify an EmbeddedDocumentExtractor.class or a Parser.class, a ParsingEmbeddedDocumentExtractor will be automatically added with an AutoDetectParser. Before Tika 1.15, if a user failed to pass in an EmbeddedDocumentExtractor or a Parser, Tika would skip embedded files.

2. XMLParsing – Users can send in their own XMLReader (StAX), SAXParser (SAX), SAXParserFactory (SAX) or DocumentBuilder (DOM). Parsers that use XML parsing will use these resources for XML parsing.

3. PasswordProvider – If you know the password to password protected files, you can send in a PasswordProvider via the ParseContext.

4. ExecutorService – For parsers that use an ExecutorService, users can pass in their own ExecutorService.

Parser Specific

1. HtmlParser

2. TesseractOcrParser

3. PDFParser

4. Microsoft Parser (as of Tika 1.15)