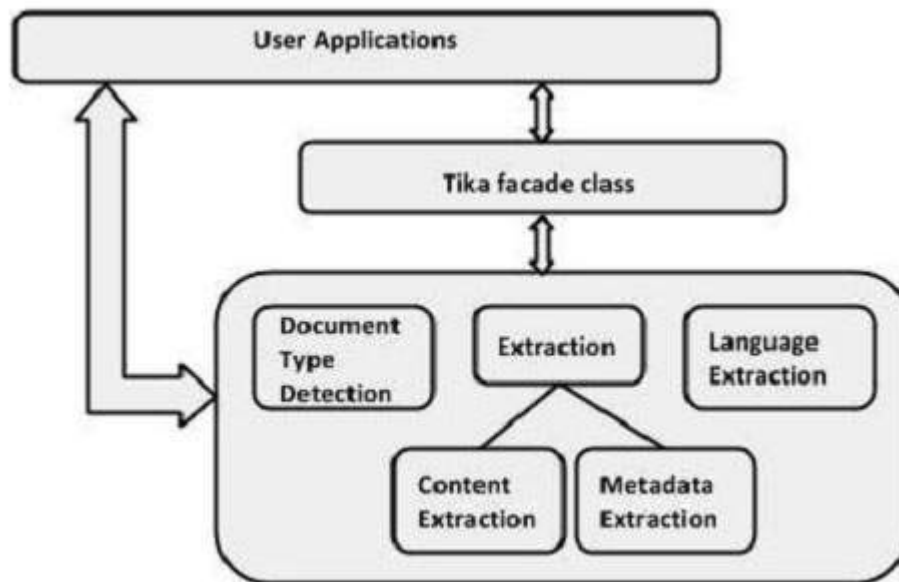


TIKA - REFERENCED API

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

Copyright © tutorialspoint.com

Users can embed Tika in their applications using the Tika facade class. It has methods to explore all the functionalities of Tika. Since it is a facade class, Tika abstracts the complexity behind its functions. In addition to this, users can also use the various classes of Tika in their applications.



Tika Class *facade*

This is the most prominent class of the Tika library and follows the facade design pattern. Therefore, it abstracts all the internal implementations and provides simple methods to access the Tika functionalities. The following table lists the constructors of this class along with their descriptions.

package : org.apache.tika

class: Tika

S.No.	Constructor and Description
1	Tika Uses default configuration and constructs the Tika class.
2	Tika <i>Detectordetector</i> Creates a Tika facade by accepting the detector instance as parameter
3	Tika <i>Detectordetector, Parserparser</i> Creates a Tika facade by accepting the detector and parser instances as parameters.
4	Tika <i>Detectordetector, Parserparser, Translatortranslator</i> Creates a Tika facade by accepting the detector, the parser, and the translator instance as parameters.

5

Tika *TikaConfigconfig*

Creates a Tika facade by accepting the object of the TikaConfig class as parameter.

Methods and Description

The following are the important methods of Tika facade class:

S.No.	Methods and Description
1	String parseToString <i>Filefile</i> This method and all its variants parses the file passed as parameter and returns the extracted text content in the String format. By default, the length of this string parameter is limited.
2	int getMaxStringLength Returns the maximum length of strings returned by the parseToString methods.
3	void setMaxStringLength <i>intmaxLength</i> Sets the maximum length of strings returned by the parseToString methods.
4	Reader parse <i>Filefile</i> This method and all its variants parses the file passed as parameter and returns the extracted text content in the form of java.io.reader object.
5	String detect <i>InputStreamstream, Metadatametadata</i> This method and all its variants accepts an InputStream object and a Metadata object as parameters, detects the type of the given document, and returns the document type name as String object. This method abstracts the detection mechanisms used by Tika.
6	String translate <i>InputStreamtext, StringtargetLanguage</i> This method and all its variants accepts the InputStream object and a String representing the language that we want our text to be translated, and translates the given text to the desired language, attempting to auto-detect the source language.

Parser Interface

This is the interface that is implemented by all the parser classes of Tika package.

package : org.apache.tika.parser

Interface : Parser

Methods and Description

The following is the important method of Tika Parser interface:

S.No.	Methods and Description
-------	-------------------------

1	parse <i>InputStreamstream, ContentHandlerhandler, Metadatametadata, ParseContextcontext</i>
---	---

This method parses the given document into a sequence of XHTML and SAX events. After parsing, it places the extracted document content in the object of the ContentHandler class and the metadata in the object of the Metadata class.

Metadata Class

This class implements various interfaces such as CreativeCommons, Geographic, HttpHeaders, Message, MSOffice, ClimateForecast, TIFF, TikaMetadataKeys, TikaMimeKeys, Serializable to support various data models. The following tables list the constructors and methods of this class along with their descriptions.

package : org.apache.tika.metadata

class : Metadata

S.No.	Constructor and description
-------	-----------------------------

1	Metadata Constructs a new, empty metadata.
---	--

S.No.	Methods and Description
-------	-------------------------

1	add <i>Propertyproperty, Stringvalue</i> Adds a metadata property/value mapping to a given document. Using this function, we can set the value to a property.
2	add <i>Stringname, Stringvalue</i> Adds a metadata property/value mapping to a given document. Using this method, we can set a new name value to the existing metadata of a document.
3	String get <i>Propertyproperty</i> Returns the value <i>ifany</i> of the metadata property given.
4	String get <i>Stringname</i> Returns the value <i>ifany</i> of the metadata name given.
5	Date getDate <i>Propertyproperty</i> Returns the value of Date metadata property.

- 6 **String[] getValues** *Propertyproperty*
Returns all the values of a metadata property.
- 7 **String[] getValues** *Stringname*
Returns all the values of a given metadata name.
- 8 **String[] names**
Returns all the names of metadata elements in a metadata object.
- 9 **set** *Propertyproperty, Datedate*
Sets the date value of the given metadata property
- 10 **set***Propertyproperty, String[]values*
Sets multiple values to a metadata property.

Language Identifier Class

This class identifies the language of the given content. The following tables list the constructors of this class along with their descriptions.

package : org.apache.tika.language

class : Language Identifier

S.No.	Constructor and description
1	LanguageIdentifier <i>LanguageProfileprofile</i> Instantiates the language identifier. Here you have to pass a LanguageProfile object as parameter.
2	LanguageIdentifier <i>Stringcontent</i> This constructor can instantiate a language identifier by passing on a String from text content

S.No.	Methods and Description
1	String getLanguage Returns the language given to the current LanguageIdentifier object.

[Previous Page](#)

[Print](#)

[PDF](#)

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