

grobid-service

manual

Overview

The project grobid-service is a RESTful service implementation for accessing the grobid system. grobid-service is an open source project under the Apache License 2.0. It comes as a war file for deploying on a web container e.g. tomcat. The project also contains the libraries of grobid-core, doing the extraction work.

Setting up grobid-service

For processing an extraction, the library grobid-core needs the trained models being contained in a special file structure. All required files organized in that structure are contained in the project grobid-home and must be downloaded separately. For referring to that folder, grobid-core uses the java system property [org.grobid.home](#). Please set this property to the grobid home folder (GROBID_HOME).

If you use tomcat as web container, you can pass system properties via setting the environment variable CATALINA_OPTS.

```
CATALINA_OPTS= -Dorg.grobid.home=GROBID_HOME
```

The grobid-service can be customized via changing the settings in a property file. The property file must have the name `grobid_service.properties` and can also be contained in GROBID_HOME. The grobid-service will resolve the file automatically. If you do not want to store that file in GROBID_HOME, you can also store to a folder of your choice, lets say GROBID_SERVICE_HOME and point the system property named [org.grobid.service.home](#) to that folder. If such a property file is not contained in GROBID_HOME, the system will check the folder GROBID_SERVICE_HOME to detect that file.

Customize grobid-service

You can customize the grobid-service by changing the properties in file `grobid_service.properties`. A list of all properties is given here:

Name of property	Description
org.grobid.service.admin.pw	The password of the admin page of the grobid-service instance

Use of grobid-service

Table 1 shows the provided resources corresponding to the HTTP verbs, to use the grobid-service. Imagine, that GROBID_HOST is the host on which the service runs. A syntactic description as WADL file is available under GROBID_HOST/application.wadl. Note, that each resource is available under uri GROBID_HOME/grobid/. A path containing brackets like {} means, that the string inside the brackets determines a variable.

uri	verb	request type	response type	description
GROBID_HOME/grobid/admin/{pw}	GET	-	text/html	Admin section of the grobid-service.
GROBID_HOME/grobid/processHeaderDocument	POST, PUT	multipart/form-data	application/xml+tei	Uploads the origin document which shall be extracted into TEI and extracts only the header data.
GROBID_HOME/grobid/processFulltextDocument	POST, PUT	multipart/form-data	application/xml+tei	Uploads the origin document which shall be extracted into TEI.
GROBID_HOME/grobid/processDate	POST, PUT	application/x-www-form-urlencoded	application/xml+tei	Parse a raw date and return the corresponding normalized date.
GROBID_HOME/grobid/processHeaderNames	POST, PUT	application/x-www-form-urlencoded	application/xml+tei	Parse a raw sequence of names from a header section and return the corresponding normalized authors.
GROBID_HOME/grobid/processCitationNames	POST, PUT	application/x-www-form-urlencoded	application/xml+tei	Parse a raw sequence of names from a header section and return the corresponding normalized authors.
GROBID_HOME/grobid/processAffiliations	POST, PUT	application/x-www-form-urlencoded	application/xml+tei	Parse a raw sequence of affiliations and return the corresponding normalized affiliations with address.

Table 1: resources and verbs to access the grobid-service