

GrobidQuantitiesParser

Grobid Quantities is a module of Grobid that specialised in the recognition of any expressions of measurements (e.g. pressure, temperature, etc.) in textual documents such as PDF publications.

Measurements are parsed normalised and converted into SI units.

To use its capabilities with Tika, one must install the server endpoint created for Grobid Quantities to extract measurement units from text passed to it.

Installing Grobid-quantities

The best approach is to run Grobid-quantities via docker.

TLDR: The following command will start the grobid-quantities image on port 8060 (the default port for grobid-quantities):

```
docker run -t --rm --init -p 8060:8060 lfoppiano/grobid-quantities:${latest_grobid_quantities_version}
```

The server starts by default on port number 8060 and the server can be seen running on <http://127.0.0.1:8060>.

Preparing resources for Grobid-quantities in Tika-App

The resources to be created are 2 files: *tika-config.xml* and *GrobidServer.properties* to be supplied later.

A predefined set of configuration files are available [here](#):

```
git clone https://github.com/lfoppiano/grobid-quantities-tika-parser-resources.git grobidquantities-parser-resources
```

Alternatively is possible to create the files automatically, as described below.

Manual configuration

Create Tika-config.xml

In order to use any of the [NamedEntityParser](#) implementations in Tika, the parser responsible for handling the name recognition task needs to be enabled. This can be done by creating the *tika-config.xml* file, as follows:

```
<?xml version="1.0" encoding="UTF-8"?>
<properties>
  <parsers>
    <parser class="org.apache.tika.parser.ner.NamedEntityParser">
      <mime>text/plain</mime>
      <mime>text/html</mime>
      <mime>application/xhtml+xml</mime>
    </parser>
  </parsers>
</properties>
```

Create GrobidServer.properties

It is imperative that Tika should know on what host you are running the *grobid-quantities-server*. By default, Tika will assume your server runs on port 8060.

In order to specify any other port, you must supply a *GrobidServer.properties* file.

```
grobid.server.url=http://localhost:8060
grobid.endpoint.text=/processQuantityText
```

Running Grobid Quantities with Tika

```
#set the system property to use GrobidNERecogniser class
cd grobidquantities-parser-resources
```

```
java -Dner.impl.class=org.apache.tika.parser.ner.grobid.GrobidNERecogniser -classpath .:tika-app-2.8.0.jar:tika-  
parser-nlp-package-2.8.0.jar org.apache.tika.cli.TikaCLI --config=tika-config.xml -m https://en.wikipedia.org  
/wiki/Time
```