# IQVIA ANONYMIZER

## Requerirements

- UNIX-based Operating System (CentOS, Ubuntu, RHEL, Fedora...)
- Free disk space. 50GB will be enough, but 100GB is prefered.
- RAM. 16GB will be enough, but 32GB is always prefered.

For technical questions & support please write to josejuan.martinez@iqvia.com

## DOCKER INSTALLATION

Please, install the Docker library following these simple instructions: https://docs.docker.com/engine/install/

Choose the UNIX distribution. For example, for ubuntu it will be https://docs.docker.com/engine/install/ubuntu/

### Account in docker hub

Make sure you have a docker hub account. This will allow you to download any docker software images you may need, including IQVIA's. It's easy and free. To get one, please visit https://hub.docker.com/

### Install docker library

Once you have an account in docker hub and docker installed, login using command line:

```
sudo docker login
```

Enter your username and password and you are ready to go.

### Install Apertium (for bilingual documents cat-esp)

First, pull the image from this repository:

```
sudo docker pull docker.io/josejuanmartineziqvia/apertium
```

Then, run the container:

```
sudo docker run -d --name apertium -p "2737:2737" docker.io/josejuanmartineziqvia/apertium
```

### Check installation

To check that apertium is up and running you can:

1) Use browser, just typing the following url, replacing SERVER_IP with your server ip value: `> http://{SERVER_IP}:2737/translate?langpair=cat|spa&q=Bon%20dia`

2) Using command line, in the same server where docker is running:

```
`> wget "http://0.0.0.0:2737/translate?langpair=cat|spa&q=Bon%20dia" -O response.txt`

`> cat response.txt`
```

3) Response should be:

```
{"responseData": {"translatedText": "Buenos días"}, "responseDetails": null, "responseStatus": 200}
```

### Install the anonymizer

First, pull the image from this repository:

```
sudo docker pull docker.io/josejuanmartineziqvia/anonymizer:0.1
```

Then, create a folder to store PDF files

```
sudo mkdir anonymizer_docs
```

```
cd anonymizer_docs
```

After, create two folders: input (for original PDF to anonymizer) and output (to get the results)

```
sudo mkdir input output
```

Finally, instantiate a container using the image setting the input and output folders and connecting it to Apertium translation engine. From `anonymizer_docs` folder, where you should have `./input` and `./output` folders already created in previous steps, execute:

```
sudo docker run -d --name anonymizer -p "9090:9090" --link "apertium" --expose "2737" -v "${PWD}/input:/opt/anonymizer/input" -v "${PWD}/output:/opt/anonymizer/output" docker.io/josejuanmartineziqvia/anonymizer:0.1
```

# EXECUTION OF ANONYMIZATION

## Preparing PDF files

Being in `anonymizer_docs` folder created before, copy all the pdfs you need to anonymize.

```
sudo cp [path_to_example_pdf].pdf ./input/
```

## Request

You can launch the anonymization using a HTTP request, for example:

1) From your browser:

```
> http://[SERVER_IP]:9090/anonymizer/pdf/extract/translate/anonymize/bulk?from_lang=cat&to_lang=spa&host=http://apertium&port=2737
```

2) Using wget:

```
> wget "http://localhost:9090/anonymizer/pdf/extract/translate/anonymize/bulk?
from_lang=cat&to_lang=spa&host=http://apertium&port=2737" -O response.txt
```

```
> cat response.txt
```

IMPORTANT: First time it will take a while, since it will download several language models, some of the of about 1GB. But it's only the first time. To see the progress:

```
sudo docker logs -f anonymizer
```

## Response

You should see either in the browser or in response.txt, the following:

```
{'status': 'OK'}
```

... that will mean that the process has successfully finished and you have the anonymized .txt results in `output` folder.