

LIR_System:

An information retrieval system which takes court sentence files in pdf/doc format as input and returns the extracted info as .json in the files directory. The information retrieved contains;

- Ponentes: Judge Name
- Identificador: Sentence ID
- Radicado: Filing Number
- fechaVigencia: Effective date
- Salas: Rooms
- Secciones: Sections

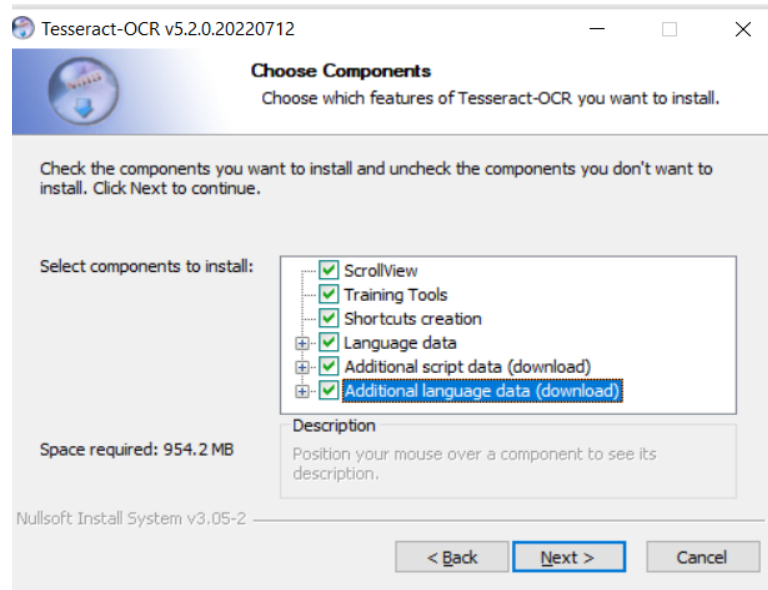
The algorithm tries to extract the most of the results but may miss some information in some cases due to multiple reasons such as broken input files(Old formats), entities with with no proper patterns etc.

Directory Tree:

```
LIR_System/
├── python3.8/
├── temp/
├── Tesseract-OCR/
├── scripts/
├── temp/
├── │
├── │   ├── file_name/
├── │   ├── file_name/
├── │   └── ...../
├── main.py
├── requirements.txt
└── readme.md
```

Installation

- Open working Directory-"./LIR_System"
- Before installing python, run "vc_redist.x64.exe" file located in './python3.9'
- Install python: .exe file located in folder './python3.9'
- Install tesseract-ocr: .exe file located in folder './Tesseract-OCR'
 - Make sure to check all boxes in select component window;



- Open command line and cd to working Directory.
 - cd\LIR_System
- Install requirements;
 - py -m pip install -r requirements
- Wait till all the requirements are installed.

Running the program:

- Open cmd and cd to working Directory
 - cd\LIR_System
- Run main.py with;
 - py main.py
- On prompt 'path to files:', paste the directory(Full path) containing sentence files.
- Wait for the program to finish processing all documents.