**Automate Election data processing**

**from PDF documents**

Table of Contents

**Installation 2**

**System requirements2**

Software package dependencies3

**Application architecture7**

Application Flow7

Formats7

File structure9

**Installation and Configuration Guide for Windows and Linux**

**Before installing and configuring the application, ensure that your system meets the hardware and software requirements mentioned below:**

**System Hardware and Software Requirements:**

**RAM: At least 16GB (48GB recommended if using GPU)**

**Storage: 20GB (or more depending on your requirements)**

**Browser: Chrome or later version**

**JavaScript: Enabled in the browser settings**

**Popup Blockers: Removed or disabled to allow application popups**

**PDF Viewer: Adobe Reader or a similar PDF viewer for viewing reports**

**Please make sure that your system meets these requirements before proceeding with the installation and configuration of the application.**

**Software on system:**

**Python 3.9 or latest**

**Dependencies:**

**anyio==3.7.1**

**astor==0.8.1**

**attrdict==2.0.1**

**attrs==21.2.0**

**Automat==20.2.0**

**Babel==2.12.1**

**bce-python-sdk==0.8.87**

**bcrypt==3.2.0**

**beautifulsoup4==4.12.2**

**blinker==1.6.2**

**blis==0.7.9**

**cachetools==5.3.1**

**catalogue==2.0.8**

**certifi==2020.6.20**

**chardet==4.0.0**

**ci-info==0.3.0**

**click==8.1.3**

**cloud-init==23.1.2**

**colorama==0.4.4**

**command-not-found==0.3**

**confection==0.1.0**

**configobj==5.0.6**

**configparser==6.0.0**

**constantly==15.1.0**

**contourpy==1.1.0**

**cryptography==3.4.8**

**cssselect==1.2.0**

**cssutils==2.7.1**

**cycler==0.11.0**

**cymem==2.0.7**

**Cython==0.29.36**

**dbus-python==1.2.18**

**decorator==5.1.1**

**distro==1.7.0**

**distro-info===1.1build1**

**ec2-hibinit-agent==1.0.0**

**et-xmlfile==1.1.0**

**etelemetry==0.3.0**

**exceptiongroup==1.1.2**

**filelock==3.12.2**

**fire==0.5.0**

**fitz==0.0.1.dev2**

**Flask==2.3.2**

**flask-babel==3.1.0**

**fonttools==4.41.0**

**future==0.18.3**

**gdown==4.7.1**

**h11==0.14.0**

**hibagent==1.0.1**

**httpcore==0.17.3**

**httplib2==0.20.2**

**httpx==0.24.1**

**hyperlink==21.0.0**

**idna==3.3**

**imageio==2.31.1**

**imgaug==0.4.0**

**importlib-metadata==4.6.4**

**incremental==21.3.0**

**isodate==0.6.1**

**itsdangerous==2.1.2**

**jeepney==0.7.1**

**Jinja2==3.1.2**

**jsonpatch==1.32**

**jsonpointer==2.0**

**jsonschema==3.2.0**

**keyring==23.5.0**

**kiwisolver==1.4.4**

**langcodes==3.3.0**

**launchpadlib==1.10.16**

**lazr.restfulclient==0.14.4**

**lazr.uri==1.0.6**

**lazy\_loader==0.3**

**lmdb==1.4.1**

**looseversion==1.3.0**

**lxml==4.9.3**

**MarkupSafe==2.1.3**

**matplotlib==3.7.2**

**more-itertools==8.10.0**

**murmurhash==1.0.9**

**netifaces==0.11.0**

**networkx==3.1**

**nibabel==5.1.0**

**nipype==1.8.6**

**numpy==1.25.0**

**oauthlib==3.2.0**

**opencv-contrib-python==4.6.0.66**

**opencv-python==4.6.0.66**

**opencv-python-headless==4.7.0.72**

**openpyxl==3.1.2**

**opt-einsum==3.3.0**

**packaging==23.1**

**paddle-bfloat==0.1.7**

**pandas==2.0.3**

**pathlib==1.0.1**

**pathy==0.10.2**

**pdf2docx==0.5.6**

**pdf2image==1.16.3**

**pexpect==4.8.0**

**Pillow==9.5.0**

**premailer==3.10.0**

**preshed==3.0.8**

**protobuf==4.23.4**

**prov==2.0.0**

**psutil==5.9.5**

**ptyprocess==0.7.0**

**pyasn1==0.4.8**

**pyasn1-modules==0.2.1**

**pyclipper==1.3.0.post4**

**pycryptodome==3.18.0**

**pydantic==1.10.9**

**pydot==1.4.2**

**PyGObject==3.42.1**

**PyHamcrest==2.0.2**

**PyJWT==2.3.0**

**PyMuPDF==1.20.2**

**pyOpenSSL==21.0.0**

**pyparsing==2.4.7**

**PyPDF2==3.0.1**

**pyrsistent==0.18.1**

**pyserial==3.5**

**PySocks==1.7.1**

**pytesseract==0.3.10**

**python-apt==2.4.0+ubuntu1**

**python-dateutil==2.8.2**

**python-debian===0.1.43ubuntu1**

**python-docx==0.8.11**

**python-magic==0.4.24**

**pytz==2023.3**

**PyWavelets==1.4.1**

**pyxnat==1.5**

**PyYAML==5.4.1**

**rapidfuzz==3.1.1**

**rarfile==4.0**

**rdflib==6.3.2**

**requests==2.25.1**

**scikit-image==0.21.0**

**scipy==1.11.1**

**SecretStorage==3.3.1**

**service-identity==18.1.0**

**shapely==2.0.1**

**simplejson==3.19.1**

**six==1.16.0**

**smart-open==6.3.0**

**sniffio==1.3.0**

**sos==4.4**

**soupsieve==2.4.1**

**spacy==3.5.4**

**spacy-legacy==3.0.12**

**spacy-loggers==1.0.4**

**srsly==2.4.6**

**ssh-import-id==5.11**

**systemd-python==234**

**termcolor==2.3.0**

**tesseract==0.1.3**

**thinc==8.1.10**

**tifffile==2023.7.10**

**tqdm==4.65.0**

**traits==6.3.2**

**Twisted==22.1.0**

**typer==0.9.0**

**typing\_extensions==4.7.0**

**tzdata==2023.3**

**ubuntu-advantage-tools==8001**

**ufw==0.36.1**

**unattended-upgrades==0.1**

**urllib3==1.26.5**

**visualdl==2.5.3**

**wadllib==1.3.6**

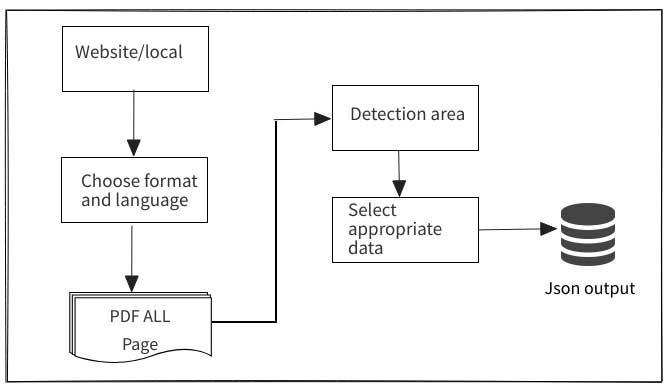
**wasabi==1.1.2**

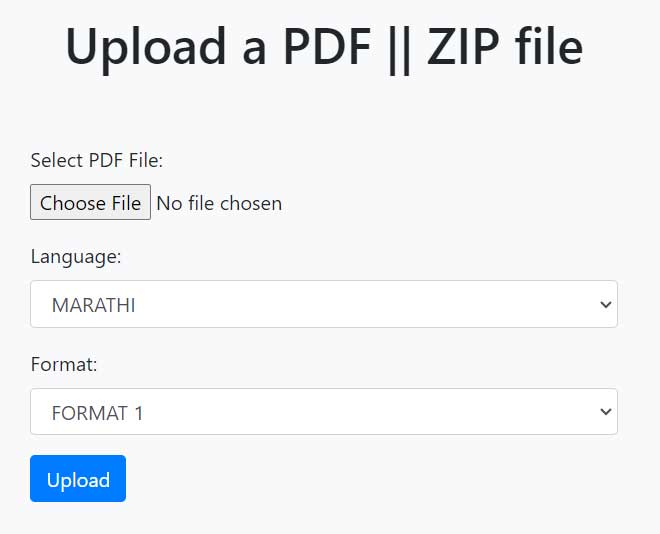
**Werkzeug==2.3.6**

**zipp==1.0.0**

**zope.interface==5.4.0**

**Application architecture**

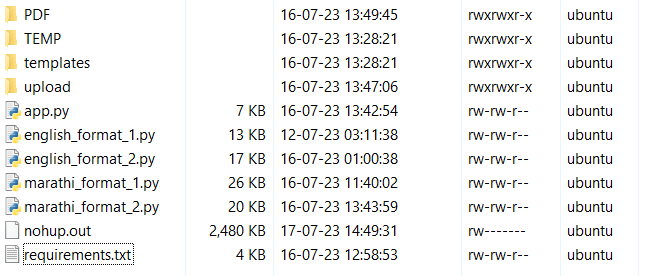
****

****

**In above image**

1. Select .pdf format with format 1,2, choose appropriate format and language than press upload.
2. Output come inside folder, the extracted text is stored in a dictionary called inputFile.json.

**File structure**

****

**PDF folder is final output for json**

**TEMP: Temp is for temporary files which will be deleted.**

**Templates: For html pages**

**Upload: save upload PDF file and convert into .pdf formats**

**App.py: Is main program, run python app.py (before install all dependencies)**

**Logical files for English 1,2 formats and Marathi 1,2 formats .py files (please don’t change inside it)**

**Nohup.out is require run script consciously on Linux server (test and developed on linux server)**

**nohup flask run --host=0.0.0.0 --port=3030 &**

**Requirements.txt having all dependencies inside to run this**

**pip install -r requirements.txt**

**kill flask on Linux**

**pkill -f flask**

**Install flask (this inside requirments.txt)**

**pip install Flask**

**Run on windows**

**python -m flask run**