

CV info extraction task

For such tasks we can solve it many methods, I thought firstly that we need to use traditional NLP solutions with regex to check my code capabilities, but after that I found that there are many better solutions because the traditional methods will not be fit for all the documents and there will be many errors with low quality so I found that I need to use better solutions.

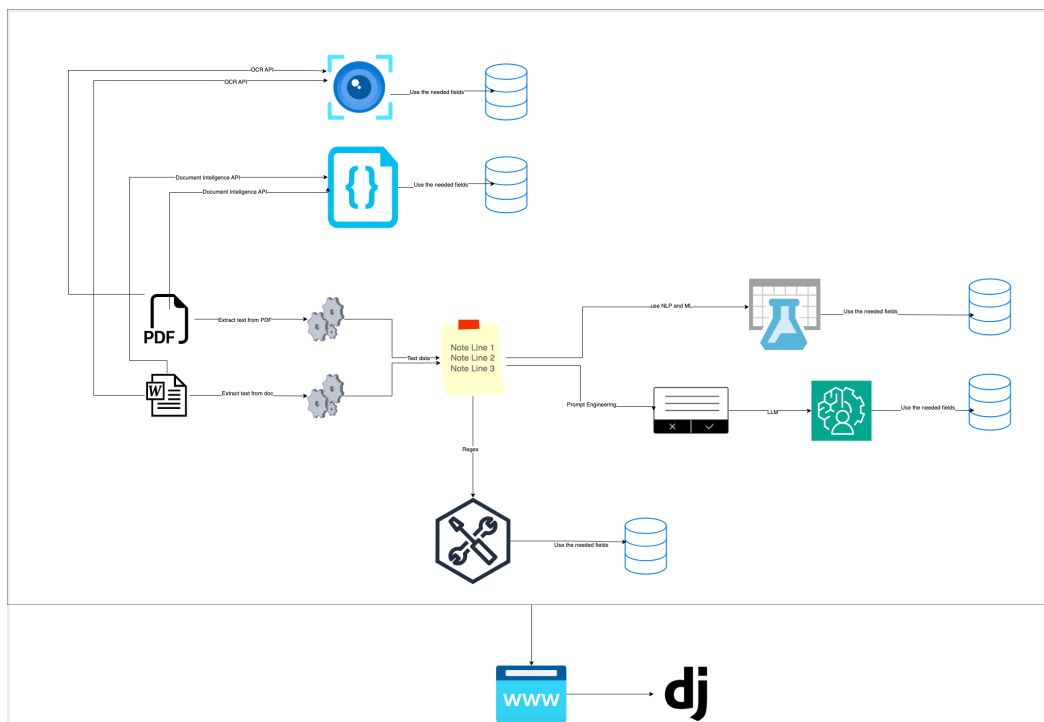
Steps to solve this task:

Some solutions:

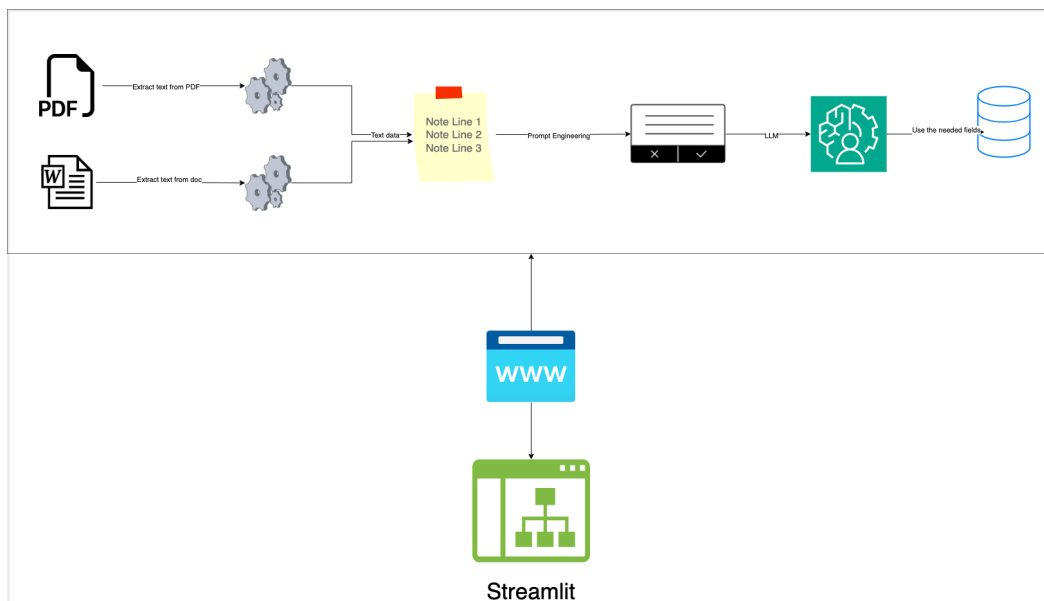
1. Traditional solutions to extract the data from each type of files, then try to search inside the text about the needed information which will may lead us to very poor quality.
 1. To use this method we need:
 1. Extract the data from both pdf and word files, we can do this using python libraries such as docx for word documents, and PyPDF2 for PDF we need to review the documentation for this library and extract the data from files, for PDF its easy task but for word, it's hard.
 2. Search about the needed data and return it.
 3. deploy the work into web app to have the UI, because we are using Python, we can use Flask, Django(better for large applications), Streamlit(I prefer it, it's real-time with best UI), or we can use it inside JAVASPRING, MERN and .NET.
 2. Extract the needed data from the files then use some NLP with ML solutions to use some NLP librairaies like Spacy, NLTK and others to search inside the data and extract the needed informations.
 1. To use this method we need:
 1. Extract the data from both pdf and word files, we can do this using python libraries such as docx for word documents, and PyPDF2 for PDF we need to review the documentation for this library and extract the data from files, for PDF its easy task but for word, it's hard.
 2. Search about the needed data using NLP and ML and return it.
 3. deploy the work into web app to have the UI, because we are using Python, we can use Flask, Django(better for large applications), Streamlit(I prefer it, it's real-time with best UI), or we can use it inside JAVASPRING, MERN and .NET.
 3. Extract the needed data from the files then use some LLM API's with prompt Engineering, for example we can use openAi API for GPT to use some models like "gpt-4o-mini" which will cost us some cost but its feasible due to the good quality for using it, we can use some free models such as llama models, HuggingFace models and others, but I decided to use the GPT one, which will lead us to excellent quality with capable cost.

1. To use this methode we need:
 1. Extract the data from both pdf and word files, we can do this using python libraries such as docx for word documents, and PyPDF2 for PDF we need to review the documentation for this libaray and extract the data from files, for PDF its easy task but for word, it's hard.
 2. Create prompt and send the extracted data with it to the api, wait for the results and use it.
 3. deploy the work into web app to have the UI, because we are using Python, we can use Flask, Django(better for large applications), Streamlit(I prefer it, it's real-time with best UI), or we can use it inside JAVASPRING, MERN and .NET.
4. Use OCR API's, theres many ready api's to do this task using OCR, such as Microsoft Azure Computer Vision OCR API, and others, which will lead us to excellent quality with high cost.
 1. To use this methode we need:
 1. Configure the API and send it the document, wait for the results and use it.
 2. deploy the work into web app to have the UI, because we are using Python, we can use Flask, Django(better for large applications), Streamlit(I prefer it, it's real-time with best UI), or we can use it inside JAVASPRING, MERN and .NET.
5. Use Document Intelligence API's, theres many ready api's to do this task, such as Microsoft Document Intelligence API, and others, which will lead us to excellent quality with high cost.
 1. To use this methode we need:
 1. Configure the API and send it the document, wait for the results and use it.
 2. deploy the work into web app to have the UI, because we are using Python, we can use Flask, Django(better for large applications), Streamlit(I prefer it, it's real-time with best UI), or we can use it inside JAVASPRING, MERN and .NET.

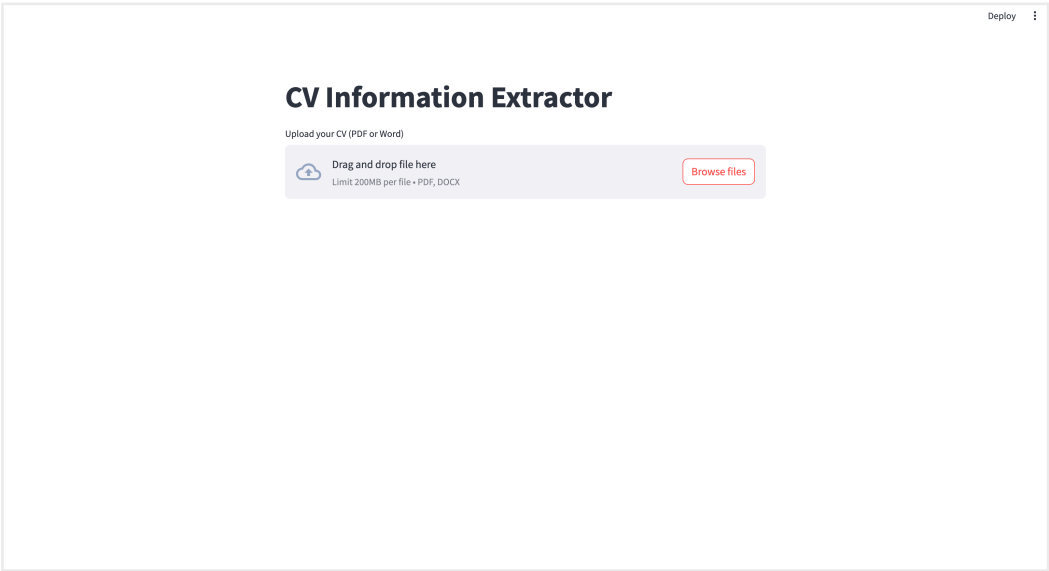
The figure below shows these all methods:



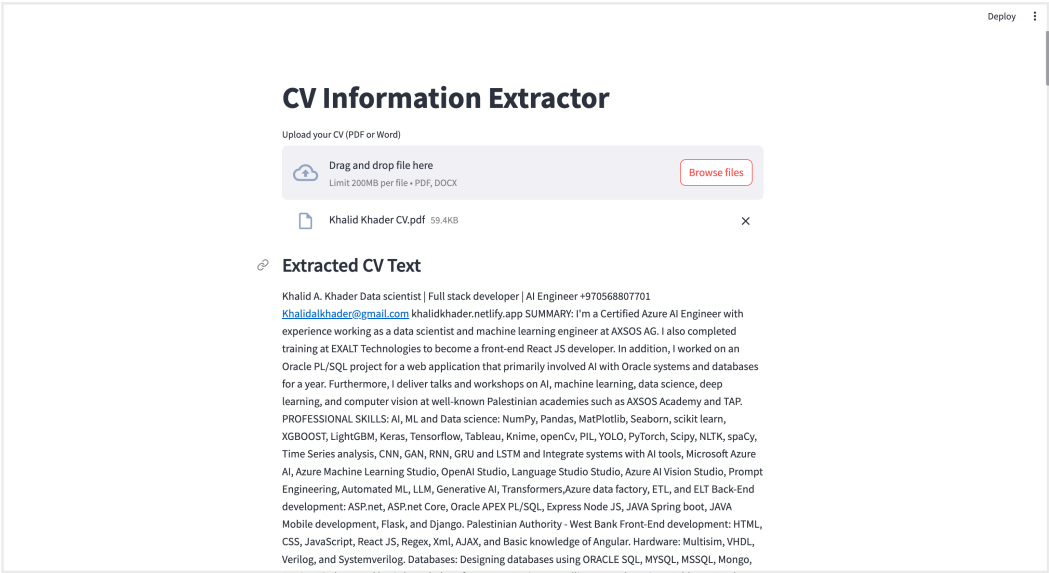
My decision was to use the 3rd way which is prompt engineering duem to the excellent quality, and capable cost for the API



Interface for Streamlit:



For PDF document:



Extracted Information

Here are the extracted details from the CV:

- **Name:** Khalid A. Khader
- **Gender:** Male
- **Age:** 23 (Date of birth: 10th May 2000)
- **Education:**
 - **Degree:** Bachelor's Degree in Computer Systems Engineering
 - **Institution:** Arab American University
 - **GPA:** 3.67 (very good)
 - **Start Date:** September 2018
 - **End Date:** July 2022
 - **Degree:** Master's Degree in Computer Engineering
 - **Institution:** Birzeit University
 - **GPA:** 4.0 (current semester)
 - **Start Date:** November 2022
 - **End Date:** Present
 - **Course:** Deep Learning using TensorFlow
 - **Institution:** OpenCV University - USA
 - **Start Date:** March 2024
 - **End Date:** (Not specified)
 - **Course:** Image Processing and Computer Vision using Python
 - **Institution:** OpenCV University - USA
 - **Start Date:** April 2024
 - **End Date:** (Not specified)
 - **Course:** Online courses
 - **Institution:** Coding Dojo - Part of Colorado University USA
 - **Start Date:** November 2022
 - **End Date:** Present
- **Work Experience:**
 - **Company:** AXSOS AG / AXSOS Academy GMBH
 - **Role:** Data Science and Machine Learning Engineer / TA
 - **Start Date:** June 2023
 - **End Date:** Present
 - **Description:** Responsible for developing the initial data science and machine learning program, specializing in computer vision, deep learning, and natural language processing. Managed curriculum creation and worked on AI projects.
 - **Company:** EXALT Technologies
 - **Role:** Frontend Development Intern
 - **Start Date:** June 2022
 - **End Date:** September 2022
 - **Description:** Learned HTML, CSS, JavaScript, and Firebase database. Created applications using React JS and various web development technologies.
 - **Company:** EXALT Technologies
 - **Role:** Design and Verification Intern
 - **Start Date:** May 2022
 - **End Date:** June 2022
 - **Description:** Learned basic concepts of Verilog programming language and System Verilog.
 - **Company:** Arab American University
 - **Role:** Team Lead in Smart Advising System
 - **Start Date:** September 2021
 - **End Date:** June 2022
 - **Description:** Created a system using APEX Oracle and implemented AI techniques to assist computer science engineering students.
 - **Company:** AXSOS AG / AXSOS Academy GMBH
 - **Role:** Full Stack Developer / TA
 - **Start Date:** November 2022
 - **End Date:** Present
 - **Description:** Excelled in code review and project management, contributed to the development of various AI and web applications.
 - **Company:** KSA Market
 - **Role:** Junior AI Consultant
 - **Start Date:** February 2024
 - **End Date:** May 2024
 - **Description:** Delivered AI services and tools for projects in different domains.
 - **Company:** TAP
 - **Role:** Professional Mentor
 - **Start Date:** September 2023
 - **End Date:** June 2024
 - **Description:** Delivered AI sessions and mentoring for the community.

For Word(Not the same CV it's old one doesn't contains the birthday):

CV Information Extractor

Upload your CV (PDF or Word)



Drag and drop file here

Limit 200MB per file • PDF, DOCX

Browse files



Khalid Khader T.docx 175.0KB



Extracted CV Text

Programming Python, C++, C, C#, JavaScript, JAVA, PL/SQL, and Assembly. Frontend development HTML, CSS, JavaScript, React JS, Regex, Xml, AJAX, and Basic knowledge of Angular. Backend development ASP.net, ASP.net Core, Oracle APEX PL/SQL, Express Node JS, JAVA Spring boot, JAVA Mobile development, Flask, and Django. Database ORACLE SQL, MYSQL, MSSQL, Mongo, SQLite, Firebase and basic knowledge of Postgres. Hardware Maltisim, VHDL, Verilog, and System Verilog. Artificial Intelligence and Data Science: NumPy, Pandas, Matplotlib, Seaborn, scikit learn, XGBOOST, LightGBM, Keras, Tensorflow, , Tableau, Work on Brain Tumors Detection using open CV and Tensorflow with Keras and PyTorch. Others Git, GitHub, Bitbucket, Docker, Linux, MacOS, JIRA, Networking, Postman, Google Collab, and AutoCAD.

master of computer engineering birzeit university. 2022 - 2024

Bachelor of computer systems engineering

arab american university 2018-2022 (4 years instead of 5 with very good)

Head of scientific community of IEEE Khalid Khader

Data science and Machine learning engineer/ta • axsos ag JULY 2023 – present Start working on data science and machine learning with basic knowledge of NumPy, Pandas, Matplotlib, Seaborn, Folium, scikit learn, XGBOOST, LightGBM, Keras, Tensorflow, SHAP, LIMA, Stats Model, Tableau Work on Brain Tumors Detection using open CV and Tensorflow with Keras and PyTorch and working in Advanced Machine learning using some Deep learning ANN like CNN and RNN. PROFISSIONAL MENTOR • TAP OCT 2023 – present

Start as a professional mentor in software development TAP program.

Teaching Assistant • axsos ACADEMY GmbH

NOV 2022 – PRESENT Full Stack Teaching Assistant. Assisting students with problem solving, different stacks, code review. Certified in Coding Dojo in MERN, Python (Django and Flask) stacks, and Java Spring boot stack. I worked a bit on development of the ALGOS LABS project for AXSOS Academy with a team of specialists in React TS, JAVA SPRING, Mongo, and Postgres database, and working on AI Sessions with some basic AI Apps.

Intern • Exalt technologies july 2022 – Sep 2022 Frontend development Html, CSS and JavaScript, firebase database, git, GitHub, AJAX, jQuery regex for validation and React JS, and create many applications with JSX, React State, React Events, Reducers, JSON web token, Hooks, React Styling, Fragments, Portals, Http Requests, Connecting Database, React Router, Single-Page Applications SPA, Deploying React Apps, Authentication, React Forms, Context API, Rest API, and some APIs.

Intern Design and verification Learned some basic concepts about Verilog programming language and System Verilog.

Smart advising system • arab american university sep 2021 - july 2022 Our System Created in APEX ORACLE, Using Artificial Intelligence Techniques. Our System aims to guide our CSE students of all student's educational levels. It will show the courses on the student's plan and the description of all courses. Check it here



Extracting information with GPT...

[courses: check it here](#)

Extracted Information

Here are the extracted details from the CV:

- **Name:** Khalid Khader
- **Gender:** Not explicitly mentioned
- **Age:** Not explicitly mentioned
- **Education:**
 - **Degree:** Master of Computer Engineering
 - **Institution:** Birzeit University
 - **GPA:** Not mentioned
 - **Start Date:** 2022
 - **End Date:** 2024
 - **Degree:** Bachelor of Computer Systems Engineering
 - **Institution:** Arab American University
 - **GPA:** Very Good (4 years instead of 5)
 - **Start Date:** 2018
 - **End Date:** 2022
- **Work Experience:**
 - **Company:** Axos AG
 - **Role:** Data Science and Machine Learning Engineer/TA
 - **Start Date:** July 2023
 - **End Date:** Present
 - **Description:** Working on data science and machine learning with basic knowledge of NumPy, Pandas, Matplotlib, Seaborn, Folium, scikit learn, XGBOOST, LightGBM, Keras, Tensorflow, SHAP, LIMA, Stats Model, Tableau. Work on Brain Tumors Detection using OpenCV and TensorFlow with Keras and PyTorch, and working in Advanced Machine Learning using some Deep Learning ANN like CNN and RNN.
 - **Company:** TAP
 - **Role:** Professional Mentor
 - **Start Date:** October 2023
 - **End Date:** Present
 - **Description:** Start as a professional mentor in software development TAP program.
 - **Company:** Axos Academy GmbH
 - **Role:** Teaching Assistant
 - **Start Date:** November 2022
 - **End Date:** Present
 - **Description:** Full Stack Teaching Assistant. Assisting students with problem solving, different stacks, code review. Certified in Coding Dojo in MERN, Python (Django and Flask) stacks, and Java Spring Boot stack. Worked on the development of the ALGOS LABS project for AXSOS Academy with a team of specialists in React TS, JAVA SPRING, Mongo, and Postgres database, and working on AI Sessions with some basic AI Apps.
 - **Company:** Exalt Technologies
 - **Role:** Intern
 - **Start Date:** July 2022
 - **End Date:** September 2022
 - **Description:** Frontend development using HTML, CSS, JavaScript, Firebase database, Git, GitHub, AJAX, jQuery regex for validation and React JS, creating many applications with JSX, React State, React Events, Reducers, JSON web token, Hooks, React Styling, Fragments, Portals, HTTP Requests, Connecting Database, React Router, Single-Page Applications (SPA), Deploying React Apps, Authentication, React Forms, Context API, REST API, and some APIs.
 - **Company:** Arab American University
 - **Role:** Intern Design and Verification
 - **Start Date:** Not mentioned
 - **End Date:** Not mentioned
 - **Description:** Learned some basic concepts about Verilog programming language and System Verilog.
 - **Project:** Smart Advising System
 - **Institution:** Arab American University
 - **Start Date:** September 2021
 - **End Date:** July 2022
 - **Description:** Created in APEX ORACLE, using Artificial Intelligence techniques to guide CSE students of all educational levels. It shows the courses on the student's plan and the description of all courses.