

AI Product Development Intern position

Screening assignment


Use case description:

- Construction management GmbH (CM) needs to process and review several construction documents for each of its projects under execution.
- After the review, they need to extract and store that information digitally and in a structured manner.
- The documents do not follow a standardized format, but contain the same five sections of information:
 - Identification of the contracting entity
 - Identification of the project, order, and services
 - Declaration of completion and compliance
 - Identification of the contractor company
 - Identification of the contractor representative
- Each section can contain multiple fields, like Tax ID., address, name, and date, among others. Some documents contain more information than others.
- We would like you to extract this information from a set of provided PDF files and present in digital format in a structured manner.

Document structure:

- Figure 1 shows an example of the document structure.
- Section 1 shows an example of the basic contracting entity identification, containing the name and address. This section can also contain contact information and a Tax ID.
- Section 2 shows an example of the project, order, and services declaration, showcasing that this information might be split into different parts of the document. The minimum information is the order ID, contract dates, service (or specialty) and specific services contracted.
- Section 3 shows an example of the declaration text, used to state that the contract took place in time and complying to requirements.
- Section 4 showcases the company information (the hired contractor), identified by their name, address, Tax ID, and address.
- Section 5 shows the signing representative's information, including at least their name, position, and the company name.

Clear Skies Contractors Ltd.



CONTRACTORS BUILDING A BRIGHTER TOMORROW

Construction Management GmbH.
 Gartenstrasse 6, 80638 Munich 1

Order N°	20230307-19A001-002.v3.	2
Project ID	19A0015672B	
Project Name	CAU Faculty Building N°3	
Project Address	Tizianstrasse 7, 80637, Munich, Bavaria	

Specialist contractor declaration:

As a specialist contractor, we declare that the contracted services were carried out in time accordance with the requirements. 3

Installation date:
October 2020 to June 2022.

Service:
Facade work 2

Specific services:
Supply and installation of curtain walls and windows.

This declaration is made on behalf of Clear Skies Contractors Ltd.

Company information:
 Clear Skies Contractors Ltd.
 Tax ID. 20/999/512315
 Braunstrasse 81, Munich, Bavaria 4
 Tel. 0444-666 888-0
 Fax. 0444-666 888-20
 contact@clearskies.com

Marc Uwe
 Managing director
 Clear Skies Contractors Ltd. 5
 TAX ID. 21/867/543926
 UST.ID. DE274234742

Figure 1. Example of "specialist contractor declaration" document.

Implementation assignment and expected outcomes:

- Extracting the text information.
- Detecting the key words and/or key sections.
- Structuring the extracted information and presenting it in an organized manner.
- Presenting working code in a live demo.

Conceptual assignment and expected discussion:

- Our full solution concept is using the extraction of structured information and keywords to provide an approval or rejection recommendation to the user.
- Approval would require three criteria:
 - The document is the one expected.
 - The document contains the five required sections.
 - Each section contains the required minimum information.
 - There are no incongruences in the information.
- We would like to discuss with you a potential solution concept, technologies, and architectural approach to employ to solve it.
- You can review these resources which we believe might be useful; and include them in your solution scope if you consider it fit:
 - Examples of resume information extraction:
 - Part 1: <https://medium.com/vitrox-publication/resume-information-extraction-part-1-290d561a15f4>
 - Part 2: <https://medium.com/vitrox-publication/resume-information-extraction-part-2-5732899e10db>
 - OCR Free document understanding transformer:
 - Research paper: <https://arxiv.org/pdf/2111.15664.pdf>
 - Repository: <https://github.com/clovaai/donut>