Backround information

Our customers need to find & extract information from non-machine-readable PDF's. This means they can not search the document for keywords but have to read through a document to find what they are looking for. We want to give them the ability to search for keywords in a document.

Your Task

To build a web application that displays a document and allows the users to:
1. Scroll through all the pages of the document
2. Enter keywords and search the document for them
3. Highlight words that match their keywords

Imagine, we have already done the OCR on the document, so we provided you with two sample files:
1. tokens.json (OCR extracted Tokens)
2. images.zip (11 extracted Images)

Every token has a unique ID as a key and a nested key-value-repairs containing the word, page and x1, y1, x2, y2-coordinates.

The images are in PNG-format with the dimensions 2481×3509.

Limitations
1. The data (images, tokens, search terms and results) has to be transmitted by a Python-Flask backend to the Front-End.
2. Please use vue.js for the Front-End
3. The application should be able to handle different loads of data (amounts of pages & sizes of images).
4. Limit yourself to a max. of 3 hours. Be aware that the time might not be sufficient, so you might need to prioritize.
5. We value functionality over design – so don't worry about the looks.