

How to Use the Modularized CV_Handler Project

1. Overview

The modularized version of CV_Handler separates key functionalities into dedicated Python modules. This makes the code cleaner, easier to maintain, test, and reuse.

2. Project Structure

```
src/  
__init__.py      # Initializes the package  
uploader.py      # Functions to upload documents to Supabase  
parser.py        # Functions to extract text from PDF and DOCX  
sorter.py        # Functions to sort documents by title  
searcher.py      # Functions to search keywords in documents  
classifier.py     # Functions to classify documents  
stats.py         # Functions to show statistics about documents
```

3. How to Use in Python

You can import and use each module in your main script or Jupyter Notebook like this:

```
from src.parser import extract_text_from_pdf  
from src.searcher import search_documents  
from src.sorter import sort_documents_by_title  
from src.classifier import classify_documents  
from src.stats import get_document_stats
```

4. Example Usage

Here is an example of how you might use these in a script:

```
# Load and parse documents
```

How to Use the Modularized CV_Handler Project

```
docs = load_documents_from_supabase()
parsed_docs = [extract_text(doc) for doc in docs]

# Sort by title
sorted_docs = sort_documents_by_title(parsed_docs)

# Search for keywords
matched = search_documents(sorted_docs, 'machine learning')

# Classify
labeled_docs = classify_documents(matched)

# Get stats
get_document_stats(labeled_docs)
```

5. Tips

- Keep each module focused on a single responsibility.
- Write unit tests for each module.
- Use `__init__.py` to expose common functions if needed.
- You can replace notebook cells with scripts for production.