**Documentation for extract\_tables.py**

This documentation provides a brief overview of how to run the extract\_tables.py script, change parameters, and install necessary dependencies.

**Description**

extract\_tables.py is a Python script designed to extract tables from a PDF file and save the details in a TXT file. The script uses a pre-trained model for table detection.

**Dependencies**

Before running the script, you need to install the following dependencies:

* torch
* transformers
* pdf2image
* tqdm

You can install these dependencies using pip:

pip install torch transformers pdf2image tqdm

Additionally, the script uses the microsoft/table-transformer-detection model from the Hugging Face Transformers library. The model will be automatically downloaded when the script is run for the first time.

**Usage**

To run the script, use the following command:

* python detect\_tables.py <pdf\_path> <threshold> <output\_txt\_file> [--batch\_size BATCH\_SIZE][--upper\_threshold UPPER\_THRESHOLD]
* <pdf\_path>: Path to the input PDF file.
* <threshold>: Confidence threshold for table detection (e.g., 0.9).
* <output\_txt\_file>: Path to the output TXT file.
* [--batch\_size BATCH\_SIZE]: (Optional) Number of pages to process in a batch. Default is 5.
* [--upper\_threshold UPPER\_THRESHOLD]: (Optional) Upper confidence threshold for table detection. Default is 1.0.

**Example**

python detect\_tables.py C:\path\to\your.pdf 0.9 C:\path\to\output.txt --batch\_size 10 --upper\_threshold 0.95

This command will process C:\path\to\your.pdf with a confidence threshold range of 0.9 to 0.95, saving the results to C:\path\to\output.txt, and using a batch size of 10 pages at a time.

**Getting Help**

To view the help message and see the available options, run:

python detect\_tables.py --help

**Summary**

1. **Install dependencies**:

pip install torch transformers pdf2image tqdm

1. **Run the script**:

python detect\_tables.py <pdf\_path> <lower\_threshold> <output\_txt\_file> [--batch\_size BATCH\_SIZE] [--upper\_threshold UPPER\_THRESHOLD]

1. **View help**:

python detect\_tables.py --help

This documentation provides a comprehensive guide on how to run the extract\_tables.py script, change parameters, and install necessary dependencies.