

# Python Script for Automatic Image Sorting by Text Recognition

Dr. Yannic Kerkhoff<sup>1,2</sup>

<sup>1</sup>Freie Universität Berlin, CRC1449 “Dynamic Hydrogels at Biointerfaces”

<sup>2</sup>Zuse Institute Berlin, IT and Data Services Group

---

## Prerequisites

- Python 3.6 or higher
- Required packages:
- `pip install easyocr pillow`
- Note: EasyOCR will automatically install PyTorch and other dependencies.

## How to Use

- Run the script in your Python environment (e.g. Thonny IDE, VSC, ...)
- When prompted, select the folder containing your images
- Select the destination folder for matching images
- Enter the text string you want to search for in the images
- The script will process each image and copy those containing your search text

## How It Works

- The script attempts to open each file in the input folder as an image
- For valid images, it performs OCR (Optical Character Recognition)
- If the search string is found in the recognized text, the image is copied to the output folder
- Progress and results are displayed in the console

## Notes

- The tool supports all image formats that PIL (Python Imaging Library) can open
- OCR is performed using both German and English language models
- Text matching is case-insensitive
- Processing time depends on image size and quantity

## Troubleshooting

- If no text is recognized, try with clearer images or different search terms
- For better recognition, ensure images have good contrast and resolution
- If the script seems slow, processing large images with OCR is resource-intensive