Technical challenge

Python developer

The challenge

Implement a sorting algorithm in a jupyter notebook that sorts words from left to right line by line.

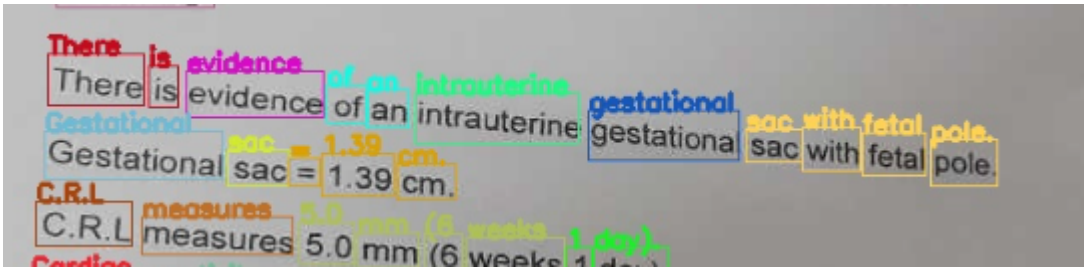
The dataset to work with can be downloaded from the following link:
https://drive.google.com/file/d/1R5D-pUqEYv4ZWywS-cOvw1uyxzfy-sVt/view?usp=sharing

In the zip file we have the source file with a json, each json looks like this:


```
{
    "ocr": [
        {
            "points": [
                [
                    2788.39599609375,
                    2310.83642578125
                ],
                [
                    2883.11865234375,
                    2310.83642578125
                ],
                [
                    2883.11865234375,
                    2346.357177734375
                ],
                [
                    2788.39599609375,
                    2346.357177734375
                ]
            ],
            "segmentationScore": 0.9469254444525644,
            "text": "cents",
            "textScore": 0.999970555305481,
            "score": 0.9468975625221202
        }
    ]
}
```

The “ocr” object contains a list of zones on the image that needs to be sorted in the order mentioned in the challenge.

The final deliverable version of the notebook should be able to:

- Load an image and json path
- Apply sorting on the zones
- Draw them on the image and plot in the notebook
- Sorted zones will have a different color based on what line they appear, example:
- 

- Should be able to sort a document with ~300 words in <= 0.5 seconds

Delivering the code

1. The code should be delivered using a Github repository.

2. Please commit and push your code often. We are interested to see the process of how you’ve built your solution.

Timeline

- The challenge should be completed between 6-20 hours.
- If you can finish it in a single day, that is a bonus.