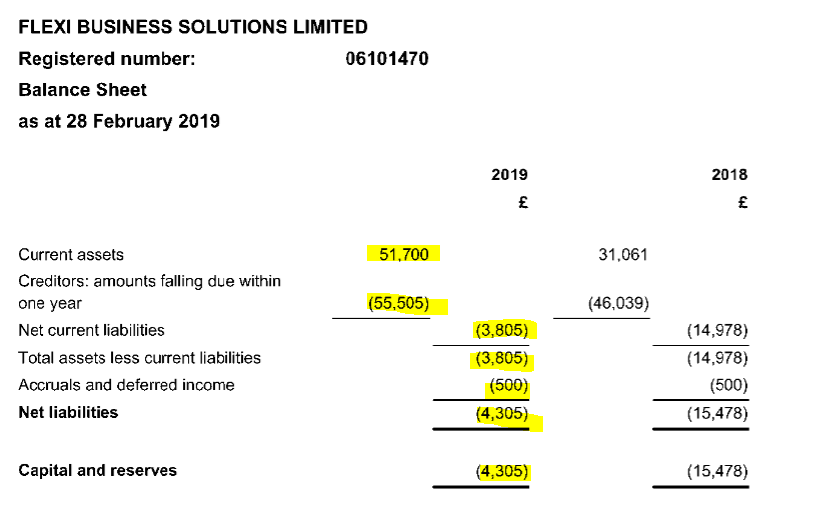
**Problem Statement Focus:** To check for NLP and Python capability to extract data elements from the financial statement documents.

**Data to be provided –**

* **Training & Development Set -** 500 Text files (OCR output of images)

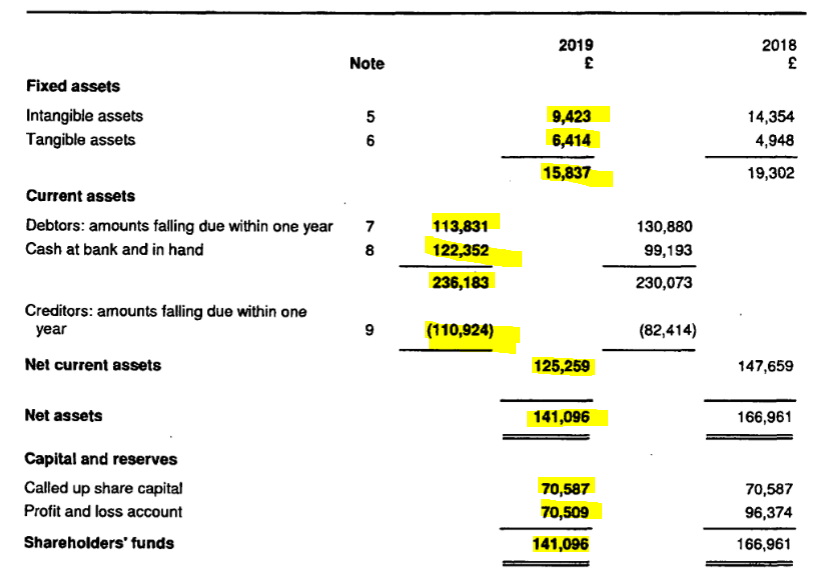
**Sample Data to be Extracted**

Highlighted text on the image is required to be extracted



Fields to be extracted are:

* Current Assets : 51,700
* Creditors: amounts falling due within one year : 55,505
* Net current liabilities : (3,805)
* Total assets less current liabilities : (3,805)
* Accruals and deferred income : (500)
* Net liabilities : (4,305)
* Capital and reserves : (4,305)



Fields to be extracted are:

* Fixed asset :  15837
* Intangible assets : 9423
* Tangible assets : 6414
* Current Assets : 236,183
* Debtors: amounts falling due within one year : 113,831
* Cash at bank and in hand : 122,352
* Creditors: amounts falling due within one year : (110,924)
* Net current assets : 125,259
* Net assets : 141,096
* Capital and reserves: -
* Called up share capital : 70,587
* Profit and loss accounts 70,509
* Shareholders funds : 141,096

Please note, data only for year 2019 is to be extracted.