

Canopy Technical Assignment

Given a single-page PDF, extract the table and output it into Excel format.

See the page below for a screenshot of the table in the PDF and an example of how it should look in Excel on output.

The top row of the Excel file should be the headers of the table. Each row in the table should correspond to one Excel row, and likewise for Excel columns.

How you decide to identify the table is entirely up to you. Some ideas include but are not limited to:

- A custom config file indicating where each table is (string-based matching, column distinction, etc.)
- Using OpenCV for table detection or row separation
- A third-party library such as Camelot-Py or Tabula-Py
- Hardcoded strings

To streamline the process, it is recommended to use Poppler to extract text from the PDF. For instance, the following command line invocation of the Poppler pdftotext utility will output a file with each word's location on the page, removing any text watermarks in the process:

```
pdftotext -f 1 -l 1 -r 300 -nodiag -bbox-layout  
"canopy_technical_test_input.pdf" "canopy_technical_test_output.html"
```

Poppler:

<https://poppler.freedesktop.org/>

Poppler can also be used in a Conda environment (especially relevant for Windows developers):
<https://anaconda.org/conda-forge/poppler>

Bonuses, but not required:

- If any columns contain dates, then standardize all entries in YYYY/MM/DD format
- If any columns contain numbers, then standardize all entries in ###.## format (removing commas; see below Excel screenshot for an example)

Note that the above asks are only bonuses. Ultimately, preference will be given to clean, modularized code.

Input:

Booking Date	Txn Date	Booking Text	Value Date	Debit	Credit	Balance
31.03.2018	31.03.2018	Initial Balance				0.00
01.04.2018	01.04.2018	VALUE DATED BALANCE BROUGHT FORWARD	01.04.2018		180,431,640.00	180,431,640.00
06.04.2018	06.04.2018	INTEREST-FIXED TERM LOAN Contract No: 3001-AA18091ZN72C Interest rate: 0.810000% Capital: 1,000,000,000.00 Period: 16.03.2018 - 06.04.2018 Days: 21/360	06.04.2018	472,500.00		179,959,140.00
06.04.2018	06.04.2018	INTEREST-FIXED TERM LOAN Contract No: 3001-AA1809166QNF Interest rate: 0.810000% Capital: 1,000,000,000.00 Period: 23.03.2018 - 06.04.2018 Days: 14/360	06.04.2018	315,000.00		179,644,140.00
06.04.2018	06.04.2018	FOREX SPOT EUR/JPY 130.7271	06.04.2018		472,500.00	180,116,640.00
06.04.2018	06.04.2018	FOREX SPOT EUR/JPY 130.7021	06.04.2018		315,000.00	180,431,640.00
09.04.2018	09.04.2018	INTEREST-FIXED TERM LOAN Contract No: 3001-AA18091NTT8N Interest rate: 0.810000% Capital: 500,000,000.00 Period: 26.03.2018 - 09.04.2018 Days: 14/360	09.04.2018	157,500.00		180,274,140.00
09.04.2018	09.04.2018	FOREX SPOT EUR/JPY 131.1407	09.04.2018		157,500.00	180,431,640.00
10.04.2018	10.04.2018	INTEREST-FIXED TERM LOAN Contract No: 3001-AA18091NL9XG Interest rate: 0.810000% Capital: 500,000,000.00 Period: 27.03.2018 - 10.04.2018 Days: 14/360	10.04.2018	157,500.00		180,274,140.00
10.04.2018	10.04.2018	FOREX SPOT EUR/JPY 131.1953	10.04.2018		157,500.00	180,431,640.00
		Carried forward				180,431,640.00

Output:

Booking Date	Txn Date	Booking Text	Value Date	Debit	Credit	Balance
2018/03/31	2018/03/31	Initial Balance				0.00
2018/04/01	2018/04/01	VALUE DATED BALANCE BROUGHT FORWARD	2018/04/01		180431640	180431640
		INTEREST-FIXED TERM LOAN Contract No: 3001-AA18091ZN72C Interest rate: 0.810000% Capital: 1,000,000,000.00 Period: 16.03.2018 - 06.04.2018 Days: 21/360				
2018/04/06	2018/04/06		2018/04/06	472500		179959140
		INTEREST-FIXED TERM LOAN Contract No: 3001-AA1809166QNF Interest rate: 0.810000% Capital: 1,000,000,000.00 Period: 23.03.2018 - 06.04.2018 Days: 14/360				
2018/04/06	2018/04/06		2018/04/06	315000		179644140
		FOREX SPOT EUR/JPY 130.7271				
2018/04/06	2018/04/06		2018/04/06		472500	180116640
		FOREX SPOT EUR/JPY 130.7021				
2018/04/06	2018/04/06		2018/04/06		315000	180431640
		INTEREST-FIXED TERM LOAN Contract No: 3001-AA18091NTT8N Interest rate: 0.810000% Capital: 500,000,000.00 Period: 26.03.2018 - 09.04.2018 Days: 14/360				
2018/04/09	2018/04/09		2018/04/09	157500		180274140
		FOREX SPOT EUR/JPY 131.1407				
2018/04/09	2018/04/09		2018/04/09		157500	180431640
		INTEREST-FIXED TERM LOAN Contract No: 3001-AA18091NL9XG Interest rate: 0.810000% Capital: 500,000,000.00 Period: 27.03.2018 - 10.04.2018 Days: 14/360				
2018/04/10	2018/04/10		2018/04/10	157500		180274140
		FOREX SPOT EUR/JPY 131.1953				
2018/04/10	2018/04/10		2018/04/10		157500	180431640
		Carried forward				180431640