

# Rechnungsdaten Extraktion

# Programmieraufgabe: Rechnungsdaten Extraktion

## Aufgabe

Implementiere ein Programm, das PDF- und TXT-Dateien als Input verarbeitet.

Ziel ist es, spezifische Informationen aus der Datei zu extrahieren und diese in einem JSON-Format darzustellen. Das JSON-Objekt soll folgende Struktur haben:



i

```
BBD¥Æ¼æəƐƐf³ə¥DbBDi jkSTUDN
BBD¥Æ¼æəƐƐf³¼°F£ßDbBDRSPRSPSaYRDN
BBDƐÆf°ßF³½ÆF¥°Ɛ³½¥£FF$DbBSSaPRRN
BBDƐÆf°ßF³½ÆF¥°Ɛ³əÆFF$DbBSRRPRR
```

.

## ⊠ Anforderungen

### Input

- Eine Datei im PDF- oder TXT-Format.
- Der Dateiname wird als Kommandozeilenargument übergeben.

### Output

- Ein JSON-Objekt mit den folgenden Feldern:
- rechnungs\_nr: Die Rechnungsnummer als String, wie sie auf der Rechnung erscheint.
- rechnungs\_datum: Das Rechnungsdatum, wie es auf der Rechnung steht. Bonus: Im Format TT.MM.JJJJ.
- gesamt\_betrag\_brutto: Der Gesamtbetrag (Brutto) , wie er auf der Rechnung steht. Bonus: Als € Dezimalzahl.
- gesamt\_betrag\_netto: Der Gesamtbetrag (Netto), wie er auf der Rechnung steht. Bonus: Als € Dezimalzahl.

### Regeln

- Das Ergebnis muss in Form eines Script/Programm auf NodeJS-Basis eingereicht werden.
- Node-Alternativen (TSNode, Bun, Deno, ...) sind erlaubt.
- Es darf alles verwendet werden: externe NPM-Module, Libraries, APIs, Binaries, Docker-Container, etc...

- **Ausnahme:** Es darf keine API verwendet werden, die explizit "Rechnungen auslesen" als Funktion anbietet, wie z.B. <https://www.mindee.com/product/invoice-ocr-api>
- Das Script muss mindestens eine der Beispiel-Dateien korrekt verarbeiten.
- Das Script muss von einem OMM-Mitarbeiter ausgeführt werden können.
  - API-Keys zu ggf. verwendeten APIs müssen nicht bereitgestellt werden, wenn diese öffentlich (bezahlt oder unbezahlt) zugänglich sind.
- Du darfst uns jederzeit kontaktieren und Fragen stellen, z.B. auch ob etwas erlaubt ist.

#### Zeitaufwand

- Mit dem richtigen Lösungsansatz **kann** die Aufgabe(alles Levels) in 1-2h erfüllt werden.
- Uns ist bewusst, dass das unter den gegebenen Umständen aber unrealistisch ist und auch nicht Ziel der Aufgabe.

#### Abgabe

- Die Quellen des Programms als ZIP-Datei per E-Mail oder File-Link.
- Falls npm-Module verwendet werden, auf jeden Fall die package.json mitschicken.
- Bitte keine node\_modules oder ähnliches mitschicken.
- Falls du das Projekt auf GitHub oder GitLab veröffentlichst, reicht auch der Link.

#### Hinweise

- Es ist in Ordnung, wenn du nicht alle Level lösen kannst. Die Aufgabe ist bewusst so gestaltet, dass sie herausfordernd ist.
- Die "Level" sind aufsteigend, wenn du Level 4 löst, dann wirst du vermutlich auch Level 1-3 mit demselben Ansatz gelöst bekommen, usw.
- Es gibt keine "falsche" Lösung, solange sie die Anforderungen erfüllt.

Viel Erfolg bei der Umsetzung der Aufgabe! Wenn du Fragen hast, stehen wir gerne zur Verfügung.