



Jméno: **Jakub Michna**

Datum: **10.10.2023**

# 1 Aplikace pro obchodní zástupce prodávajícím státním organizacím

## 1.1 Zadání

Tento report informuje o semestrální práci vytvořené pro předmět BI-PYT v zimním semestru v akademickém roce 2024/25. Program usnadňuje práci obchodním zástupcům, kteří prodávají státním organizacím. Využívá se povinnosti zveřejňování smluv v Registru smluv, který vyplývá ze zákona č. 340/2015 Sb., o registru smluv, který vstoupil v účinnost 1. července 2016. Tento zákon ukládá vybraným subjektům, jako jsou orgány státní správy, územní samosprávné celky, veřejné instituce nebo státem vlastněné podniky, povinnost zveřejnit smlouvy, jejichž hodnota přesahuje 50 000 Kč bez DPH. Samotný program se skládá ze tří hlavních částí. První se věnuje stahování informací a smluv z registru smluv pro konkrétní dodavatele. Další část extrahuje informace o dodaném zboží, množství a cenách pomocí jazykového modelu ze stažených smluv. Finální část je webová aplikace, kde se zobrazí po vybrání odběratele jednotlivé informace o smlouvách, které odběratel uzavřel s vybranými dodavateli. Pro svou práci jsem vybral dodavatele, kteří se soustředí na dodávání textilu a zdravotnických pomůcek do sociálních služeb.

## 1.2 Použité metody a postupy

V první části jsem využil open-source knihovny Selenium, pomocí které lze simulovat uživatelské chování v prohlížeči. Pro konkrétní znalost dodavatele, lze záznamy v registru smluv snadno získat. Po zadání IČO, lze na stránce vidět konkrétní záznamy, ke kterým lze přistupovat pomocí tlačítka detail. Toto tlačítko nás přepojí na stránku s konkrétním záznamem. Jsou zde uvedené informace, jako předmět smlouvy, ID verze smlouvy, odběratel, dodavatel, a další. Také zde je libovolný počet souborů, příslušející dané smlouvě. Tyto informace, lze pomocí knihovny a přednastaveného místa stáhnout stáhnout. Na stránku s výpisy se dá dostat tlačítkem zpět. Toto tlačítko nás ale vrátí na stranu 1 a tlačítko zpět tedy není vhodné k použití, protože bychom extrahovali informace pouze z první strany. Řešení, které jsem aplikoval, spočívá v otevírání detailu smlouvy v novém okně a následném zavření okna po extrakci údajů. Extrahované textové údaje se ukládají do souboru csv.

V druhé části jsou načteny údaje pro konkrétního dodavatele ze souboru csv. Pro každý záznam se načtou všechny odpovídající soubory, a extrahuje se z nich text. Text ze souborů s příponou pdf nemusí být snadno získatelný, proto je použit postup převedení souboru na obrázky pomocí knihovny pdf2image a následné získání textu pomocí knihovny pytesseract. Extrahovaný text se předá jazykovému modelu spolu s požadovaným promptem. Byly použity jazykové modely gpt-3.5-turbo a gpt-4-turbo. Výstupem jazykového modelu je text ve formátu json. Extrahované informace týkající se dodaného zboží přidáme do csv souboru.

Ve webové aplikaci se po vybrání odběratele vyfiltrují údaje ze souboru csv. Jednotlivé záznamy seřazené od nejčerstvějších se vypíší pomocí knihovny streamlit.

## 1.3 Výsledky

Dosažené výsledky v podobě webové aplikace mohou sloužit jako informativní materiál pro obchodní zástupce zaměřující se na státní organizace. Navržená webová aplikace ušetří přípravu na obchodní jednání a obchodní zástupce bude u některých produktů znát objednané zboží a cenu. Některé z údajů vyextrahovaných ze stažených souborů nejsou správné. To je způsobeno kvalitou vstupních dokumentů, přesností knihovny pytesseract a použitých jazykových modelů. Pro získání přesných informací mohou obchodní zástupci použít tlačítko u jednotlivých smluv, které je automaticky přepojí na stránku se smlouvou a ověřit si informace na vlastní oči.

## 1.4 Možná pokračování SP

Úskalí aplikace je v používání jazykových modelů, které přirozeně halucinují (vymýšlí si) a jsou ve většině případů zpoplatněny. Má kombinace využití jazykových modelů vyšla na 0,5 Kč za zpracovanou smlouvu. Výzkum by šlo rozvíjet a zřejmě i zlepšovat použitím jiných způsobů pro extrakci textu ze smluv a úpravou promptu u jazykových modelů.