

Návrh systému

Táto kapitola sa zameriava na návrh informačného systému určeného na poloautomatické spracovanie historických dokumentov a šifrovacích kľúčov. Návrh systému vychádza z identifikovaných požiadaviek používateľov a z potrieb integrácie AI modulu, ktorý zabezpečuje automatizované spracovanie obrazových dát. Cieľom návrhu je vytvoriť modulárny a rozšíriteľný systém, ktorý umožňuje efektívnu spoluprácu medzi používateľom a automatizovanými nástrojmi pri analýze historických dokumentov.

1 Prípady použitia

Funkčné požiadavky systému sú zachytené prostredníctvom use case diagramov, ktoré ilustrujú interakciu jednotlivých aktérov so systémom. Diagramy znázorňujú hlavné prípady použitia a ich vzájomné vzťahy. V systéme vystupujú tria aktéri: používateľ, administrátor a AI modul.

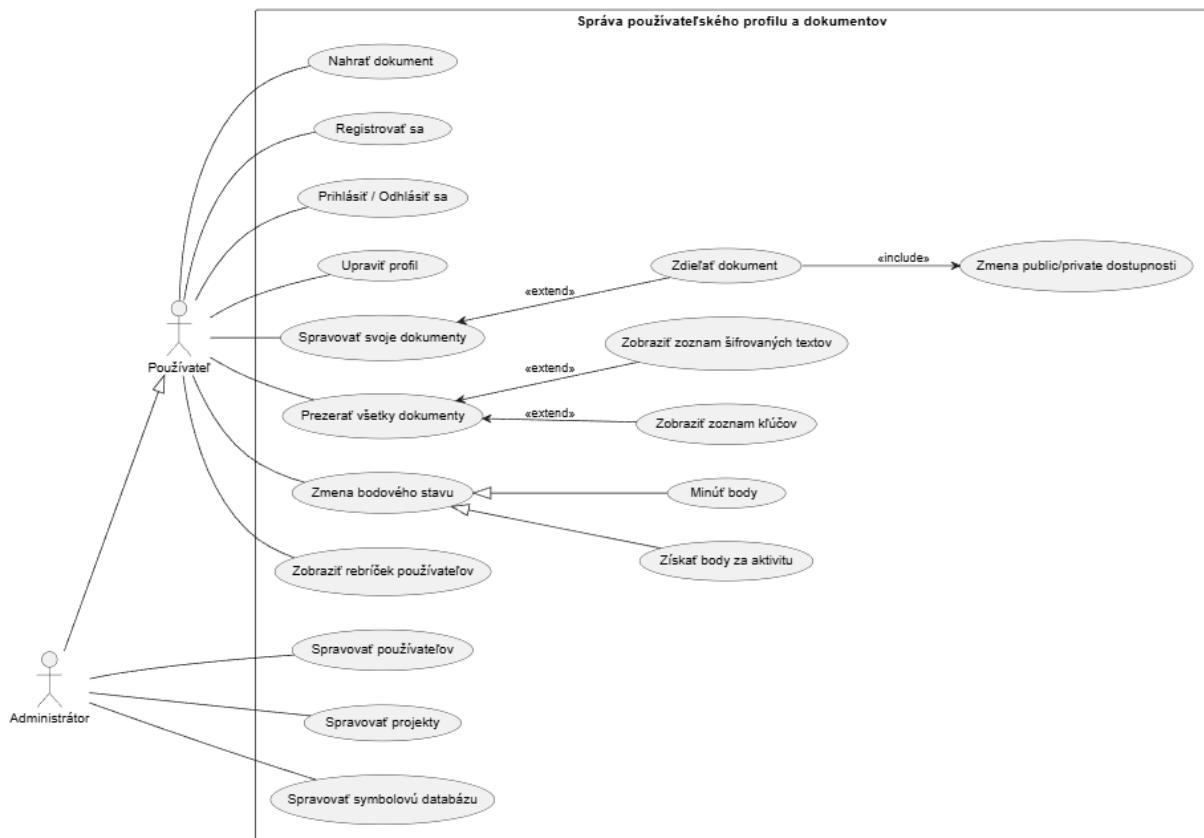
1.1 Správa používateľského profilu a dokumentov

Na obr. 1.1 je znázorený use case diagram zameraný na správu používateľského profilu a dokumentov. Hlavným aktérom je používateľ, ktorý má možnosť registrovať sa do systému a následne sa prihlásiť a odhlásiť. Po prihlásení môže používateľ upravovať svoj profil a spravovať dokumenty, ktoré do systému nahral.

Správa dokumentov zahŕňa nahrávanie dokumentov, ich prezeranie a zdieľanie s ostatnými používateľmi. Zdieľanie dokumentu predstavuje rozšírenie základného prípadu použitia správy dokumentov a zahŕňa aj nastavenie jeho verejnej alebo súkromnej dostupnosti. Používateľ má zároveň možnosť prezerať zoznam všetkých dostupných dokumentov, ako aj špecializované zoznamy šifrovaných textov a šifrovacích kľúčov.

Súčasťou systému je bodový mechanizmus, ktorý slúži ako motivačný prvok. Používateľ môže získať body za aktivitu v systéme a míňať ich na vybrané funkcionality.

Administrátor vystupuje ako aktér s rozšírenými oprávneniami. Má prístup k prípadom použitia zameraným na správu používateľských účtov, projektov a symbolovej databázy, ktorá je využívaná pri spracovaní historických dokumentov.



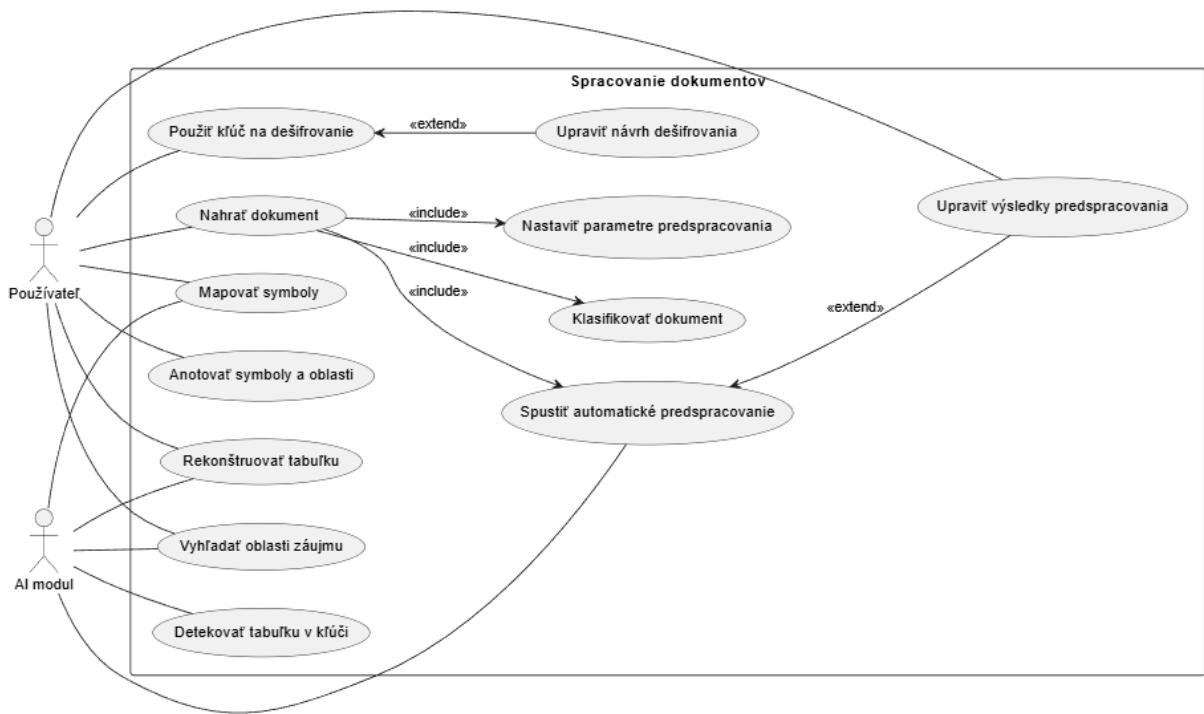
1.2 Spracovanie dokumentov

Druhý use case diagram, znázornený na obr. 1.2, sa zameriava na samotný proces spracovania dokumentov. Hlavným aktérom je opäť používateľ, ktorý iniciauje spracovanie nahraním dokumentu do systému.

Počas nahrávania má používateľ možnosť nastaviť parametre predspracovania a spustiť automatické spracovanie dokumentu. Tento proces zahŕňa klasifikáciu dokumentu a aplikáciu AI modulov, ktoré detegujú oblasti záujmu, tabuľkové štruktúry a jednotlivé symboly.

AI modul v tomto prípade vystupuje ako podporný aktér, ktorý zabezpečuje automatizované úlohy, ako je detekcia tabuľiek v šifrovacích klúčoch, vyhľadávanie oblastí záujmu a návrh rekonštrukcie tabuľkových štruktúr. Výsledky automatického spracovania môže používateľ manuálne upravovať, čím sa zabezpečuje poloautomatický charakter systému.

Používateľ má zároveň možnosť manuálne anotovať symboly a oblasti, mapovať symboly na ich významy a rekonštruovať tabuľky šifrovacích klúčov. V záverečnej fáze môže byť existujúci šifrovací klúč použitý na dešifrovanie textu, pričom návrh dešifrovania je možné iteratívne upravovať.



2 Tok dát v systéme

Tok dát v systéme je znázornený pomocou diagramu toku dát (Data Flow Diagram), ktorý ilustruje interakciu medzi používateľom, aplikačnou vrstvou, backendom, AI modulmi a dátovým úložiskom. Diagram zachytáva hlavné procesy spracovania dokumentov a tok informácií medzi jednotlivými komponentami systému.

Proces začína autentifikáciou používateľa prostredníctvom prihlásovacieho mechanizmu. Po úspešnom prihlásení môže používateľ inicovať nahranie dokumentu do systému. Dokument je prijatý aplikačnou vrstvou a prostredníctvom zabezpečeného REST API odoslaný backendu systému.

Backend vystupuje ako centrálny koordinačný prvok, ktorý zabezpečuje spracovanie požiadaviek používateľa. Po prijatí dokumentu backend uloží vstupné dátá do dátového úložiska a následne odovzdá dokument na spracovanie AI modulom. Komunikácia s AI modulmi prebieha prostredníctvom samostatného REST API rozhrania, čo umožňuje ich nezávislý vývoj a škálovanie.

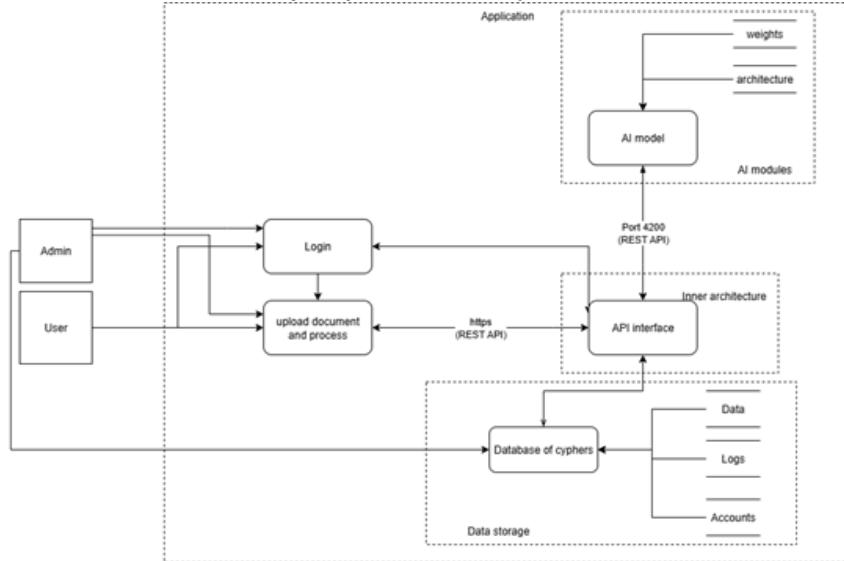
AI moduly aplikujú jednotlivé kroky spracovania dokumentov, ako je úvodná kontrola dokumentu, predspracovanie obrazu, detekcia štruktúr a symbolov a návrh anotácií. Výsledky spracovania sú po dokončení inferencie vrátené späť backendu, kde sú uložené spolu s metadátami dokumentu.

Backend následne sprístupňuje výsledky spracovania aplikačnej vrstve, ktorá ich zobrazuje používateľovi. Používateľ má možnosť výsledky manuálne upravovať, dopĺňať anotácie alebo opakovane spúšťať vybrané kroky spracovania s upravenými parametrami. Týmto spôsobom systém podporuje iteratívny a poloautomatický pracovný postup.

Dátová vrstva systému zahŕňa databázu šifier, používateľských účtov a systémových logov. Backend zabezpečuje konzistentný prístup k týmto dátam a kontroluje oprávnenia

používateľov pri jednotlivých operáciách. Logovacia vrstva slúži na zaznamenávanie priebehu spracovania dokumentov a systémových udalostí, čo umožňuje spätnú analýzu a ladenie systému.

Celkový tok dát je navrhnutý tak, aby minimalizoval závislosti medzi jednotlivými komponentami a umožnil jednoduché rozšírenie systému o nové AI moduly alebo ďalšie funkcionality bez zásahu do existujúcej architektúry.



3 Sekvenčný diagram spracovania dokumentu

Na znázornenie dynamického správania systému pri spracovaní dokumentu bol použitý sekvenčný diagram, ktorý je uvedený na obr. 3. Diagram zachytáva časovú postupnosť interakcií medzi jednotlivými komponentami systému a ilustruje úlohu backendu ako centrálneho koordinátora spracovania.

Proces začína akciou používateľa, ktorý prostredníctvom aplikačnej vrstvy nahrá dokument do systému. Frontend odosielá dokument spolu s metadátami backendu systému, ktorý zabezpečuje jeho uloženie do dátového úložiska. Po potvrdení úspešného uloženia backend iniciuje spracovanie dokumentu odoslaním požiadavky na AI modul.

AI modul vykonáva jednotlivé kroky spracovania dokumentu, ktoré zahŕňajú úvodnú kontrolu dokumentu, predspracovanie obrazu, detekciu štruktúr a symbolov a návrh anotácií. Tieto kroky prebiehajú interne v rámci AI modulu a sú zobrazené ako sekvencia vnútorných operácií. Po dokončení spracovania sú výsledky vrátené späť backendu.

Backend následne ukladá výsledky spracovania, ako sú anotácie a metadáta dokumentu, do dátového úložiska a po potvrdení uloženia ich sprístupňuje aplikačnej vrstve. Frontend zobrazí výsledky používateľovi, ktorý má možnosť vykonať manuálne úpravy, napríklad korekciu anotácií alebo mapovanie symbolov.

Upravené dáta sú opäťovne odoslané backendu, ktorý zabezpečuje ich aktualizáciu v dátovom úložisku. Po potvrdení aktualizácie backend poskytne frontendovej vrstve finálne výsledky, ktoré sú následne zobrazené používateľovi.

Sekvenčný diagram tak demonštruje poloautomatický a iteratívny charakter spracovania dokumentov, kde sa automatické výstupy AI modulov kombinujú s manuálnymi zásahmi používateľa. Zároveň zvýrazňuje oddelenie zodpovednosti medzi aplikačnou vrstvou, backendom, AI modulmi a dátovým úložiskom.

