

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE
FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

**ZÍSKAVANIE A VYUŽITIE INFORMÁCIÍ Z
VEREJNE DOSTUPNÝCH ZDROJOV S PODPOROU
OPEN SOURCE TECHNOLOGIÍ MACHINE
LEARNING A AI
TÍMOVÝ PROJEKT**

**Bc. Adam Klimko, Bc. Martina Krnáčová, Bc. Klaudia
Kytošová, Bc. Andrej Mikuš, Bc. Marek Mlynček, Bc. Zuzana
Stančíková, Bc. Ľubomíra Števuliaková**

Obsah

Úvod	1
1 Predstavenie tímu	2
1.1 Členovia tímu	2
1.1.1 Marek Mlynček	2
1.1.2 Zuzana Stančíková	2
1.1.3 Ľubomíra Števuliaková	3
1.1.4 Klaudia Kytošová	3
1.1.5 Martina Krnáčová	3
1.1.6 Andrej Mikuš	4
1.1.7 Adam Klimko	4
1.2 Rozvrh	4
1.3 Organizačná štruktúra	4
2 Špecifikácia požiadaviek	6
2.1 Cieľ	6
2.2 Funkcionálne požiadavky	6
2.3 Nefunkcionálne požiadavky	7
2.4 Roly	7
2.5 Špecifikácia aktivít	9
3 Ponuka	10
3.1 Návrh projektu	10
3.2 Plán	11
3.3 Predpokladané zdroje	11
4 Analýza problematiky	13
4.1 Prieskum existujúcich riešení na trhu	13
4.2 Analýza dát	16
4.2.1 EDA - Exploratory Data Analysis	16
5 Technológie	19
5.1 Data Processor	19
5.1.1 Google Search	19
5.1.2 Fuzzy Logic	19

5.1.3	Natural Language Processing	19
5.1.4	Machine learning	21
5.2	Backend	22
5.2.1	Laravel	23
5.2.2	Django	23
5.2.3	SpringBoot	24
5.2.4	Vyhodnotenie	25
5.3	Frontend	25
5.3.1	Angular	25
5.3.2	React	25
5.3.3	Vue	25
5.3.4	Vyhodnotenie	26
5.4	Databáza	26
5.4.1	SQL vs NoSQL	26
5.4.2	MySQL	27
5.4.3	MongoDB	27
5.4.4	Elasticsearch	28
5.4.5	Vyhodnotenie	29
6	Implementácia	30
6.1	Architektúra	30
6.1.1	REST	30
6.1.2	Model implementovanej aplikácie	32
6.2	Doménová časť	32
6.3	Autentifikácia	34
6.4	Práca so súbormi a s databázou	35
6.5	Inštancie projektov a ich management	35
6.6	Python crawler	37
6.7	Frontend	38
6.7.1	UI mock-up prototypy	39
7	Vyhodnotenie výsledkov	41
	Záver	42
	Zápisnice	43

Zoznam použitej literatúry	90
Prílohy	I
A Api dokumentácia	II

Zoznam skratiek

AI	Artificial intelligence
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface
CLI	Command Line Interface
CPU	Central Processing Unit
CRUD	Create,Read,Update,Delete
CSRF	Cross-site Request Forgery
DB	DataBase
DOM	Document Object Model
EDA	Exploratory Data Analysis
HTML	HyperText Markup Language
JDK	Java Development Kit
JEE	Java Enterprise Edition
JPQL	Java Persistence Query Language
JWT	JSON Web Token
MUI	Accessible library of React components
MVC	Model View Controller
NLP	Natural Language Processing
npm	Node.js package manager
ORM	Object Relational Mapping
PHP	Hypertext Preprocessor
POJO	Plain Old Java Object
RDBMS	Relational Database Management System
REST	Representational State Transfer
SQL	Structured Query Language
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UX	User Experience

Zoznam obrázkov a tabuliek

Obrázok 1	UML use-case diagram	8
Obrázok 2	UML diagram aktivít	9
Obrázok 3	Bloková schéma projektu	10
Obrázok 4	DiffBot Natural Language API Demo	14
Obrázok 5	ScrapeBox	15
Obrázok 6	ParseHub	15
Obrázok 7	Histogram zanalyzovaných portálov	18
Obrázok 8	Model aplikácie	32
Obrázok 9	Diagram tried	33
Obrázok 10	Entitno-relačný diagram	34
Obrázok 11	Ukážka dekódovaného tokenu z našej aplikácie (www.jwt.io) . .	34
Obrázok 12	Súborová štruktúra	36
Obrázok 13	Konfiguračný súbor	36
Obrázok 14	Graf procesu zbierania a hodnotenia článkov z Google	37
Obrázok 15	Registrácia/Prihlásenie	39
Obrázok 16	Zoznam projektov	39
Obrázok 17	Vytvorenie nového projektu	40
Tabuľka 1	Response stavové kódy	31
Tabuľka 2	Príklad skórovacieho slovníka	38

Úvod

Dnes žijeme v dobe, kedy sú informačné systémy neodlúčiteľnou súčasťou nášho života. Tieto informačné systémy zachytávajú obrovské množstvo dát, ktoré musíme vo väčšine prípadov ďalej manuálne spracovávať sami. Bez informácií, ako takých, nevieme vykonávať ďalšiu prácu/rozhodnutia. Pre vykonávanie dobrých rozhodnutí pri riadení procesov, či už v oblasti podnikania alebo v osobnom živote, potrebujeme mať prístup k dátam, v čo najväčšej kvalite a kvantite. Ako zdroj informácií v dnešnej dobe využívame internet, ale informácie, ktoré poskytuje, môžu byť zavádzajúce alebo často schované za množstvom nepotrebných a nepodstatných dát. Už len na získavanie týchto dát jednotliviec vynaloží veľké množstvo času a zdrojov, no výsledok nemusí byť vždy uspokojivý. Z vyššie uvedených dôvodov sa v dnešnej dobe rozvíjajú nové odvetvia v informatike, konkrétne Data science a Data mining - tieto odvetvia sa zameriavajú na získavanie a spracovávanie dát, ktoré sa ďalej spracovávajú pomocou Machine learning. V dnešnej dobe existujú riešenia, ktoré nám poskytujú už pretrénované modely a komponenty na uľahčenie vývoja a implementácie inteligentných informačných systémov.

Naším cieľom je navrhnúť riešenie, softvér, ktorý by vyhľadával ohlásené investičné projekty, rozširovanie výroby, budovanie nových priemyselných a rezidenčných priestorov, akvizície spoločností a podobne a uľahčil tak prácu, primárne tímov projektového financovania, v spoločnosti VÚB.

Softvér nahradí človeka pri vyhľadávaní obchodných príležitostí v online prostredí a získané dáta spracuje do pravidelného reportu.

1 Predstavenie tímu

1.1 Členovia tímu

1.1.1 Marek Mlynček

Študujem prvý ročník inžinierskeho štúdia na Slovenskej Technickej Univerzite v Bratislave, zameranie Aplikovaná informatika - bezpečnosť informačných systémov.

Moja bakalárska práca bola o automatizovanej detekcii niektorých z najčastejšie vyskytujúcich sa útokov na webové aplikácie. Celú prácu a aj zdrojové kódy je možné nájsť na: https://github.com/Logner109/WAIDS_pub. Zaujímam sa o rôzne druhy informačnej bezpečnosti, predovšetkým etický hacking a penetračné testovanie a bezpečnosť systémov a webových aplikácií.

Pracujem v Slovenskej sporiteľni na pozícii Security Specialist a venujem sa klientskej bezpečnosti, čo prioritne zahŕňa znižovanie dopadu hackerov na našich klientov aj v oblasti elektronického bankovníctva, ako aj platobných kariet (phishing, vishing a podobne). Taktiež sa venujem investigácii SIEM related udalosti a všetkým SOC témam.

V súvislosti s danou témou tímového projektu: v minulosti som sa podieľal na tvorbe HBOS modelu strojového učenia založeného na štatistických modeloch a histogramoch profilovaných dát.

V rámci štúdia som sa venoval strojovému učeniu a neurónovým sieťam na predmete Umelá inteligencia - UMINT.

1.1.2 Zuzana Stančíková

Som študentkou 1. ročníka inžinierskeho štúdia na Aplikovanej informatike, zameranie Bezpečnosť informačných systémov. Popri škole pracujem pre UniCredit Leasing na oddelení IT ako IT Analytik.

Čo sa týka skúseností s Machine learning a AI - okrajovo som sa stretla s touto problematikou v práci, kde som pomáhala pri tvorbe dátových setov pre projekty daného charakteru. Inak sa považujem za začiatočníka. Je to pre mňa výzva a zároveň možnosť naučiť sa niečo nové, veľmi využiteľné do budúcnosti.

Za veľké plus považujem absolvovanie predmetu Umelá inteligencia a v zimnom semestri máme aktuálne predmet Strojové učenie a neurónové siete, kde by sme mali taktiež dostať základy z umelej inteligencie, inteligentných systémov a strojového učenia.

1.1.3 Ľubomíra Števuliaková

Volám sa Ľubka Števuliaková a študujem na FEI Aplikovanú informatiku so zameraním MSUS. Počas bakalárskeho ročníka som nazbierala skúsenosti s tvorbou web stránok v jazykoch Java, PHP, Javascript. Pracovala som s technológiami React, SpringBoot, Hibernate, mariaDb.

V súvislosti s témou som úspešne ukončila predmet Umelá inteligencia a taktiež Programovanie 1, kde sme pracovali s Pythonom. Čo sa týka pracovných skúseností v odbore, 2 roky som pracovala v ITAPSe pre UniCreditLeasing na pozícii IT support.

1.1.4 Klaudia Kytošová

Som študentkou prvého ročníka inžinierskeho štúdia na FEI STU, odbor Aplikovaná informatika so zameraním na bezpečnosť informačných systémov. Počas bakalárskeho štúdia som nadobudla niekoľko vedomostí z programovacích jazykov Python, C, C++ a Java.

Čo sa týka skúseností s témou projektu, počas bakalárskeho štúdia som úspešne absolvovala predmet Umelá inteligencia, kde sme pracovali v Matlabe. Navyše tento rok v zimnom semestri mám zapísaný predmet Strojové učenie a neurónové siete. Taktiež popri škole pracujem pre ITAPS ako IT analytik a zároveň aj ako externý pracovník pre ČSOB Leasing na oddelení IT rozvoj.

1.1.5 Martina Krnáčová

Som študentka 1. ročníka inžinierskeho štúdia na FEI STU, odbor Aplikovaná informatika so zameraním na modelovanie a simulácia udalostných systémov.

Témou mojej bakalárskej práce boli progresívne webové aplikácie, kde som sa zaoberala výhodami a nevýhodami tejto pomerne novej technológie a možnosť týchto aplikácií nahradiť natívne aplikácie. Popri štúdiu pracujem ako IT analytik, kde som mala možnosť pracovať na viacerých projektoch, ktoré boli zamerané na webové aplikácie a služby, automatizáciu pomocou UiPath a Power Automate.

Mojou pracovnou náplňou je vypracovanie analýzy nových softvérových riešení, analýza automatizácie existujúcich procesov, vypracovanie UX&UI návrhov aplikácii. Taktiež sa venujem BPM nástrojom a pracujem na vytváraní procesov a webovej aplikácii pomocou BPM nástroja Camunda.

1.1.6 Andrej Mikuš

Študent 1. ročníka inžinierskeho štúdia na FEI STU v obore Aplikovaná informatika so zameraním na modelovanie systémových udalostí.

Popri štúdiu aktívne pracujem v obore vývoja softvérových aplikácií, zahrňujúcich web aplikácie a databázové systémy, ako aj v obore návrhu a správy počítačových sietí.

Vyššie jedného roka som sa podieľal na rôznych projektoch so zameraním na data science a machine learning vo firme PwC ako intern-consultant.

V rámci štúdia som absolvoval predmet Umelá inteligencia v rámci bakalárskeho štúdia a vo voľnom čase pracujem na vývoji webových a desktopových aplikácií.

1.1.7 Adam Klimko

Som študentom 1. ročníka inžinierskeho štúdia na odbore Aplikovaná informatika, so zameraním modelovanie a simulácia udalostných systémov. S témou umelej inteligencie som sa stretol počas bakalárskeho štúdia na predmete Umelá inteligencia. Tu som sa prvýkrát oboznámil s princípmi neurónových sietí. Aktuálne sa tiež zúčastňujem predmetu Strojové učenie a neurónové siete. Verím, že si na tomto predmete obohatím zručnosti, uplatniteľné práve pri implementácii zadania tímového projektu.

Mojou záľubou v rámci IT sveta je najmä vytváranie webových aplikácií pomocou moderných frameworkov. Práve tomuto sa venujem vo voľnom čase a pracujem na vlastných projektoch. Okrem toho pracujem v spoločnosti Siemens Healthineers na pozícii IT Support.

1.2 Rozvrh

Zavedli sme pravidelné meetingy: každú druhú stredu o 10:00 hod. v zimnom semestri a každý druhý pondelok o 13:00 hod. v letnom semestri, cez platformu MS Teams, kde sme predstavili pokroky z predchádzajúcich týždňov a určili sme si úlohy na ďalšie týždne.

1.3 Organizačná štruktúra

Pri práci na projektoch väčšieho rozsahu, určite nie je dobrou politikou úvaha - „všetci robia všetko“ a preto sme sa rozhodli rozdeliť náš tím do menších podskupín, aby sme tak zefektívnilí prácu tým, že sa každá skupina zameria na oblasť, ktorej sa bude primárne venovať:

1. Tím analytikov - analýza, dokumentácia, administratíva.

Členovia Tímu: Bc. Klaudia Kytošová, Bc. Zuzana Stančíková

2. Tím vývojárov - technická časť, implementácia.

Členovia Tímu: Bc. Adam Klimko, Bc. Andrej Mikuš, Bc. Marek Mlynček

3. Tím mix (analýza+vývoj) - podľa potreby pomoc obom tímom.

Členovia Tímu: Bc. Martina Krnáčová, Bc. Lubomíra Števuliaková

2 Špecifikácia požiadaviek

2.1 Cieľ

Motiváciou a hlavným cieľom vytvorenia systému je uľahčenie práce, primárne tímov projektového financovania v spoločnosti VÚB. Náš softvér efektívne nahradí človeka pri prehľadávaní internetu, za účelom nájdania investičných príležitostí. Zamestanci sa tak môžu naplno venovať aktivitám, na ktorých nie je schopný pracovať počítač.

Vďaka funkcionalite pridávania článkov a kľúčových slov podľa uváženia užívateľa, sa stáva náš softvér univerzálnym pomocníkom t.j. v konečnom dôsledku dokáže vyhľadať články z akejkoľvek kategórie a oblasti.

2.2 Funkcionálne požiadavky

Funkcionálna požiadavka je popis služby, ktorú musí softvér ponúkať a pomáha zachytiť zamýšľané správanie systému:

1. Prihlásením do systému sa rozumie zadanie platnej kombinácie mena a hesla. Systém porovná hash hesla uloženého v databáze s hashom aktuálne zadaného hesla. Pri zhode systém používateľa autentifikuje a zobrazí stránku. Prihláseniu prirodzene predchádza registrácia.
2. Bez prihlásenia nie je užívateľ autorizovaný na využívanie systému.
3. Každý prihlásený používateľ je oprávnený manipulovať s projektami, ktoré vytvoril len on sám.
4. Používateľ vidí prehľad svojich vytvorených projektov a môže im meniť stav: na aktívny alebo pozastavený. Taktiež môže modifikovať všetky uvedené parametre bližšie špecifikované v use-case diagrame na obr. 1. Projekt môže aj natrvalo vymazať.
5. Softvér ponúka užívateľovi predvolené portály a defaultný slovník kľúčových slov, ktoré sme získali dôslednou analýzou v prvej časti projektu. Užívateľ tak vie dostať ešte priaznivejšie výsledky.
6. Používateľ si môže vo svojom profile meniť heslo.
7. Odhlásením z aplikácie sa strácajú všetky oprávnenia, ktoré sú určené len aktuálne prihláseným používateľom.

8. Systém umožňuje odosielanie pravidelného reportu v stanovenom intervale na špecifikované e-mailové adresy daného projektu vo forme csv súboru, ktorý obsahuje id záznamu, url, odkaz na daný článok, časovú pečiatku odoslania reportu a pridelené skóre. Články v reporte sú zoradené zostupne podľa výšky prislúchajúceho ratingu.
9. Softvér je odmedzený na prehľadávanie webstránok len so slovenskou doménou.

2.3 Nefunkcionálne požiadavky

Nefunkcionálna požiadavka je špecifikácia, ktorá popisuje prevádzkové schopnosti systému a obmedzenia, ktoré zlepšujú jeho funkčnosť a využiteľnosť:

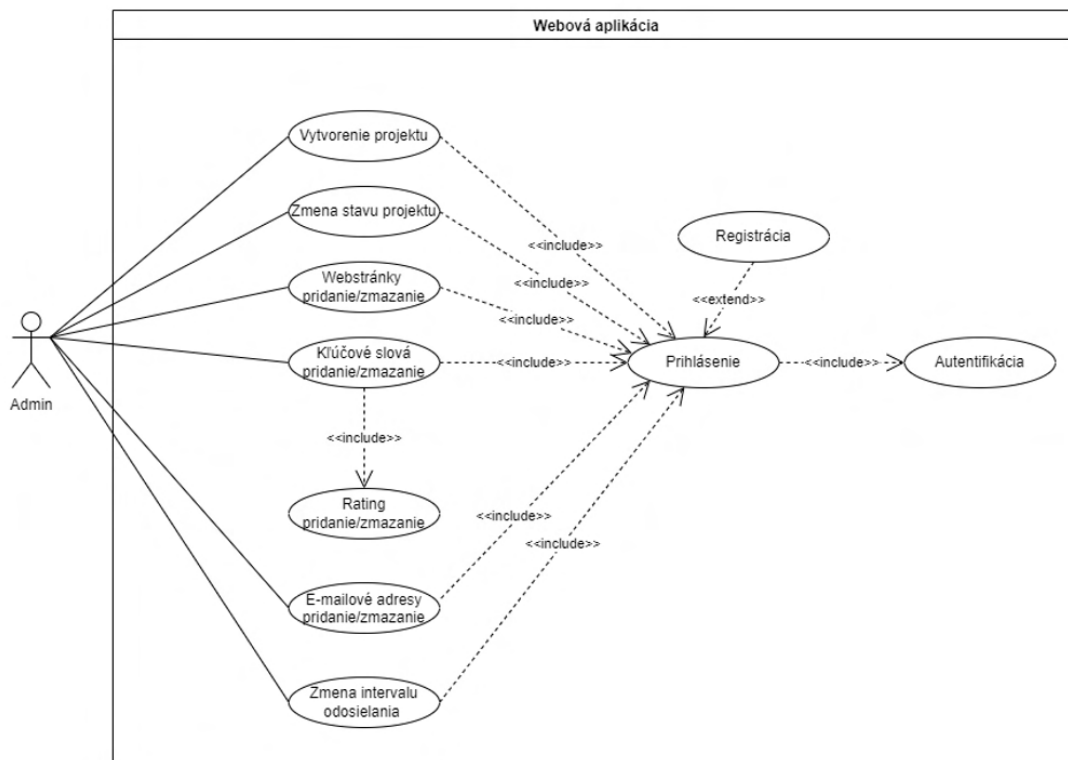
1. Vyžaduje sa prehľadný, jednoduchý systém s minimalistickým a estetickým dizajnom, ktorý je pre používateľa intuitívny.
2. Prehľadný, čitateľný a ľahko zrozumiteľný výsledný report odosielaný na zadané e-mailové adresy používateľov.
3. Spoľahlivosť a dostupnosť systému zabezpečíme dôkladným otestovaním aplikácie a teda znížením chybovosti.
4. Bezpečnosť systému docielime validáciou vstupov, autentifikáciou užívateľov pomocou JWT tokenu a ukladaním hashovaných hesiel do databázy.

2.4 Roly

Naša výsledná aplikácia bude mať len jednu rolu - Admin. Aplikácia bude navrhnutá tak, aby bola do budúcnosti rozšíriteľná, čo znamená, jednoduché pridávanie ďalších rolí.

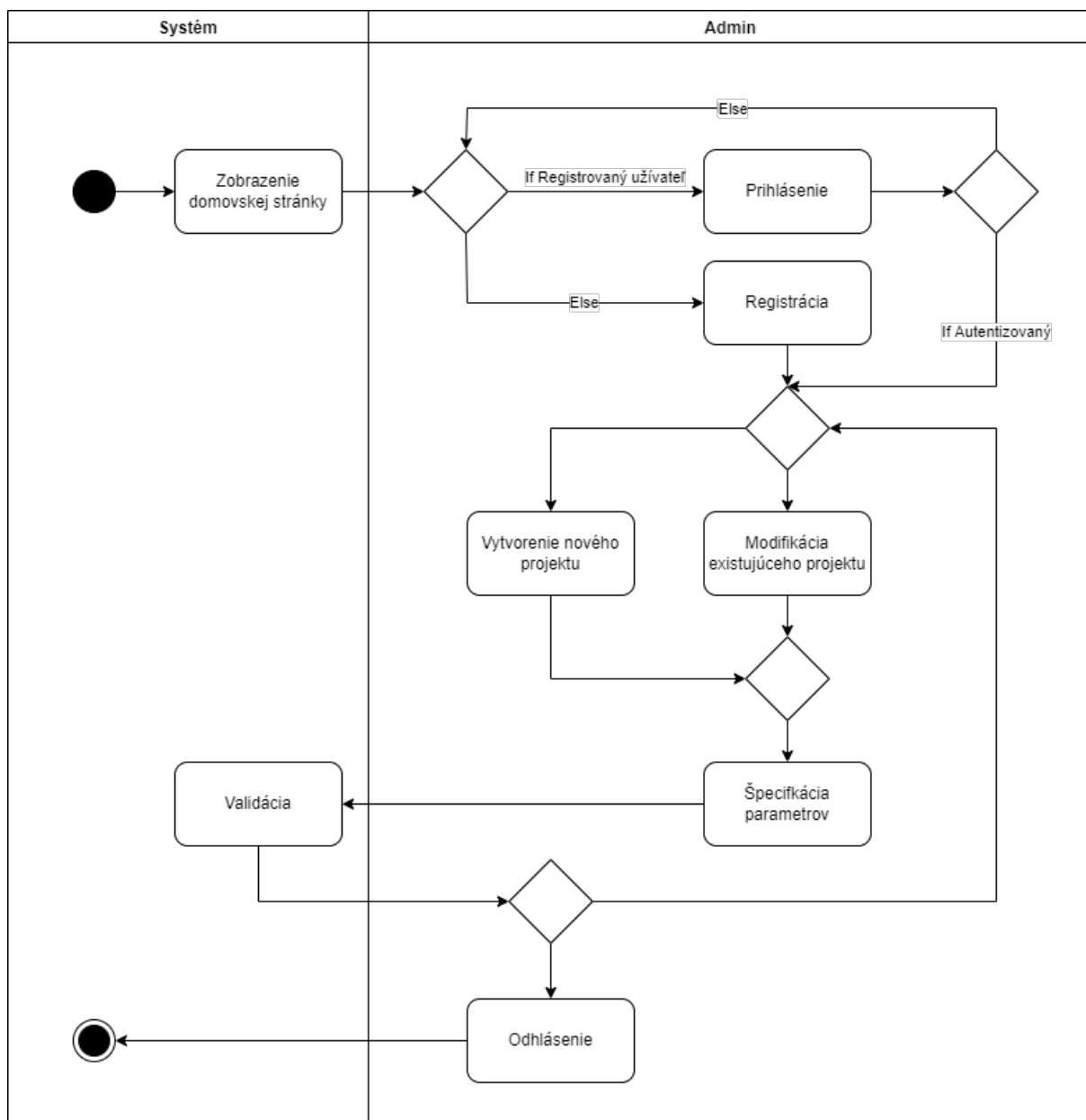
Kedže Admin, je jediná rola v aplikácii, tak po registrácii do systému je táto rola autorizovaná vykonávať všetky dostupné operácie. Pochopiteľne, pri akomkoľvek zásahu musí byť admin prihlásený a teda autentizovaný. Admin vie vytvoriť projekt a následne manipulovať s jeho aktuálnym statusom. Projekt môže prepnúť do stavu: aktívny, pozastaviť ho alebo úplne vymazať. Ďalej vie pridávať webstránky, ktoré považuje za dôveryhodné. Následne sa tieto webstránky použijú pri prehľadávaní online priestoru. Analogicky vie zakázať nedôveryhodné zdroje. Admin je oprávnený pridávať kľúčové slová a k nim prislúchajúci „rating“ tzn. dôležitosť daného slova v intervale od 0 po 100. Admin taktiež rozhoduje o tom, komu bude report doručovaný, špecifikovaním e-mailových adries užívateľov a v neposlednom rade rozhoduje o časovom intervale generovania reportu.

Pre lepšiu predstavu právomocí roly Admin sme vytvorili jednoduchý UML Use-case diagram.



Obr. 1: UML use-case diagram

2.5 Špecifikácia aktivít

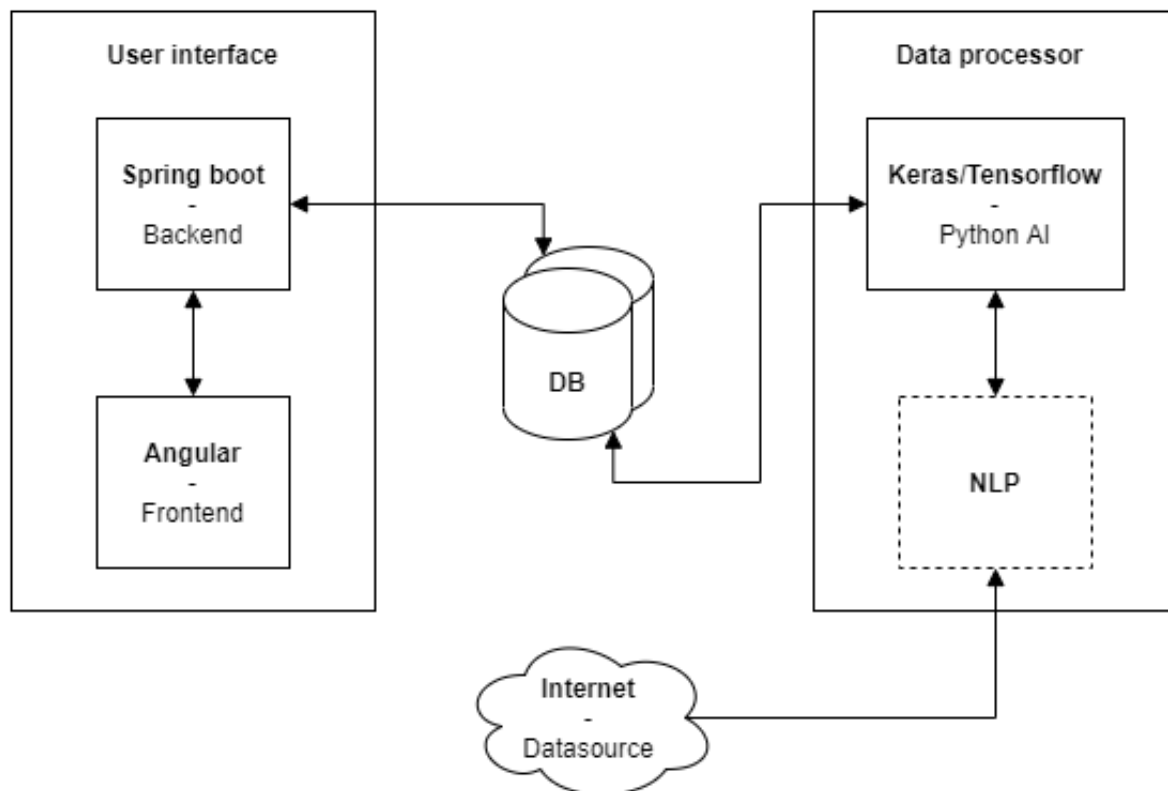


Obr. 2: UML diagram aktivít

3 Ponuka

3.1 Návrh projektu

Náš projekt sme v hrubom návrhu rozdelili na niekoľko častí. Rozdelenie môžeme vidieť na nasledujúcej blokovej schéme.



Obr. 3: Bloková schéma projektu

Aplikácia bude rozdelená na dve hlavné časti:

- **User interface:** táto časť bude slúžiť ako rozhranie pre užívateľa poskytujúca náhľad udalostí, ktoré sa dejú v ostatných častiach aplikácie a taktiež na reprezentáciu výstupných dát. Táto časť sa bude skladať z dvoch modulov:
 - **Backend** modul, ktorý bude implementovaný podľa architektúry REST, využívajúci framework SpringBoot v jazyku Java.
 - **Frontend** multiplatformový modul implementovaný ako web aplikácia za využitia frameworku React.

- **Data processor:** časť slúžiaca na získavanie surových dát, ich následnú analýzu a finálne spracovanie. Tieto spracované dáta sa budú ukladať do centrálneho úložiska. Celá táto časť bude implementovaná ako proces v jazyku Python a bude sa skladať z nasledovných modulov:
 - **NLP** modul slúžiaci na filtrovanie a predbežnú validáciu získavaných dát z internetu.
 - **Python AI** modul, ktorý bude spracovávať (kategorizovať a vyhodnocovať) dáta na základe pretrénovaného modelu.

3.2 Plán

Predbežný plán projektu sme rozdelili na nasledovné časti:

- Hlbšia analýza problematiky spolu so zadávateľom témy
- Analýza vstupných dát
- Vytvorenie tréningových dátových setov
- Tréning, nastavenie a validácia modelu
- Návrh a implementácia dátového úložiska
- Implementácia časti **Data processor**
- Detailný návrh časti **User interface** a následná implementácia
- Doladenie prepojenia všetkých častí
- Testovanie
- Nasadenie aplikácie

3.3 Predpokladané zdroje

Z dôvodu návrhu projektu bude z hľadiska nasadenia celého projektu potrebné mať k dispozícii server, ktorý za ideálnych podmienok bude disponovať:

- Operačným systémom Linux

- Kvalitnou dátovou priepustnosťou na internetovom porte
- Root/admin prístupom pre doinštalovanie potrebných aplikácií
- Aspoň 2x VCPU, 2GB RAM, úložiskom 8GB (ideálne typu SSD)

Pri vývoji nie sú požadované žiadne ďalšie špeciálne požiadavky, ktoré by boli v aktuálnom čase známe.

4 Analýza problematiky

Po úvodnej konzulácii a odprezentovaní ponuky zadávateľovi témy, Ing. Michalovi Petrášovi, nám bola odkomunikovaná predstava projektu zo strany VUB:

1. ako výstup sa vyžaduje vygenerovaný report v pravidelnom intervale so zoznamom najnovších článkov a aktualít s novými obchodnými príležitosťami,
2. vizualizácia štatistík z výstupov (počty najviac využívaných kľúčových slov v článkoch, najlepšie ohodnotené články a pod.) do grafov a dashboardov,
3. navrhnuté UI z dodanej ponuky nie je nevyhnuté. V prípade naplnenia očakávaní a zostatku času je možné doimplemetovanie UI ako nad rámec riešenia.

4.1 Prieskum existujúcich riešení na trhu

V súčasnosti už existuje niekoľko webových stránok alebo softvérov, ktoré sú založené na myšlienke web scrapingu, t.j. automatizované prehľadávanie webu, s cieľom nájsť dôležité informácie z rôznych oblastí. V tejto kapitole sa teda pozrieme na niekoľko takýchto softvérov:

- **www.diffbot.com** - webstránka, ktorá ponúka viacero nástrojov API zameraných na web scraping pomocou strojového učenia a NLP. Nástroje umožňujú extrahovať dôležité informácie z webových stránok a stránok s článkami, pričom ignorujú prvky, ktoré sa nepovažujú za jadro primárneho obsahu. Hlavnou výhodou týchto nástrojov je user-friendly použitie. To znamená, že osoba, ktorá ich používa nemusí byť technicky zdatná. Naopak veľkou nevýhodou je vysoká cena za licencie na tieto nástroje, keďže to nie je voľne dostupná platforma s nástrojmi. Ďalšou nevýhodou je, že tento nástroj nefunguje na všetkých webových stránkach. Táto webstránka so spomínanými nástrojmi je vhodná pre vývojárov a veľké technologické spoločnosti. Na obrázku nižšie môžeme vidieť krátku demo ukážku rozhrania, ktoré využíva NLP. Na ľavej strane môžeme vidieť článok, ktorý bol analyzovaný a sú v ňom vyznačené kľúčové slová. V tabuľke na pravej strane je zoznam slov, ktoré vyhodnotilo ako dôležité a tiež sú im priradené nejaké tagy.
- **ScrapeBox** - je desktopový softvér, ktorý umožňuje robiť niekoľko vecí súvisiacich s webovým prehľadávaním. Tento nástroj má viaceré funkcionality, no najznámejšie je zbieranie URL adries z viac ako 30 vyhľadávacích webov, medzi ktoré patrí Google,

Diffbot Natural Language API Demo Science by Diffbot

Facebook Buying Popular GIF Platform Giphy
In \$400 Million Deal

Select Example Clear Submit

Facebook Buying Popular GIF Platform
Giphy In \$400 Million Deal

One of Facebook's biggest acquisitions in recent memory is for a service commonly used to find GIFs of adorable puppies, Real Housewives and literal dumpster fires with broad metaphorical potential.

Facebook (FB) said Friday that it had acquired Giphy, a popular search engine for short, looping videos

Graph Entities Facts Open Facts Custom Facts Categories More

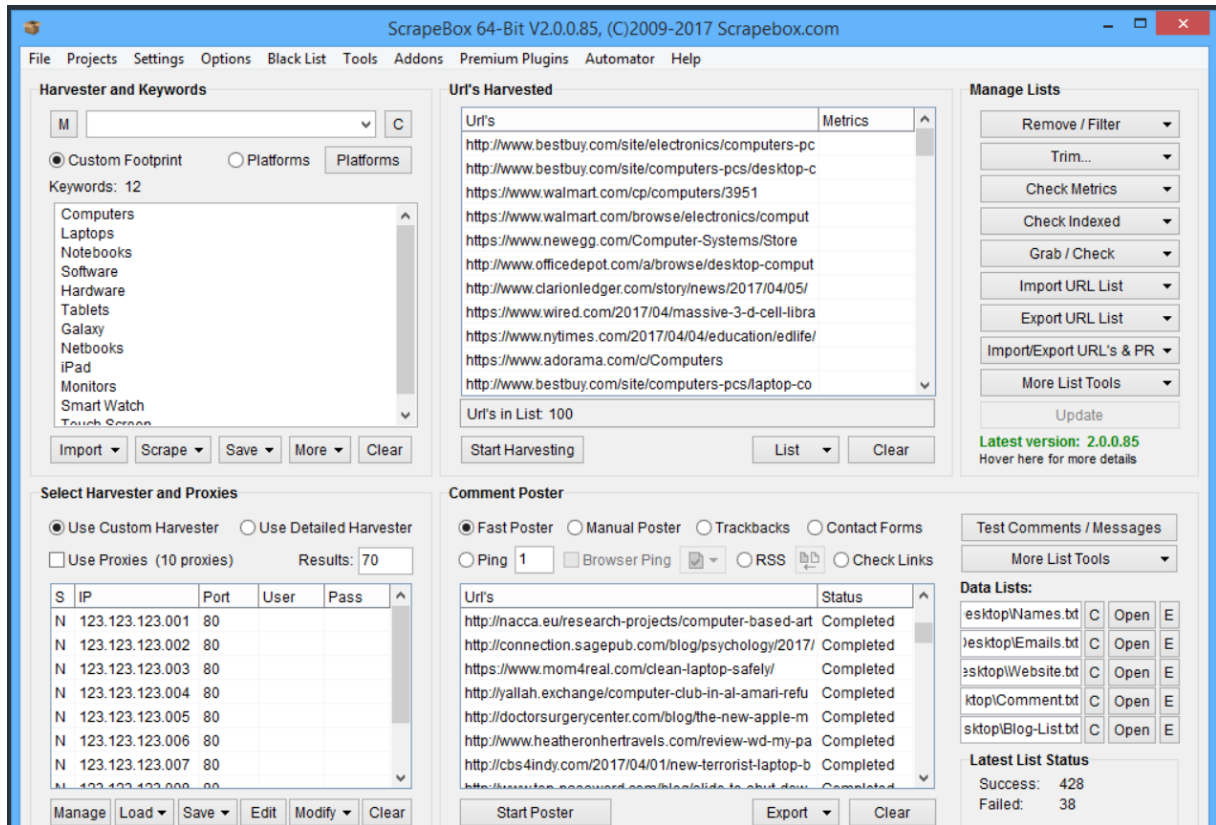
Entity	Types	Salience	Sentiment
Facebook	organization	1.00	0.00
Giphy	organization	0.98	0.77
Instagram	organization, skill, product	0.98	0.00
Graphics Interchange Format		0.95	0.52
Apple Inc.	organization, skill, product, field of work	0.88	0.00
Homer Simpson	person, fictional entity	0.84	-0.56
CNN Business	organization	0.80	0.00

Obr. 4: DiffBot Natural Language API Demo

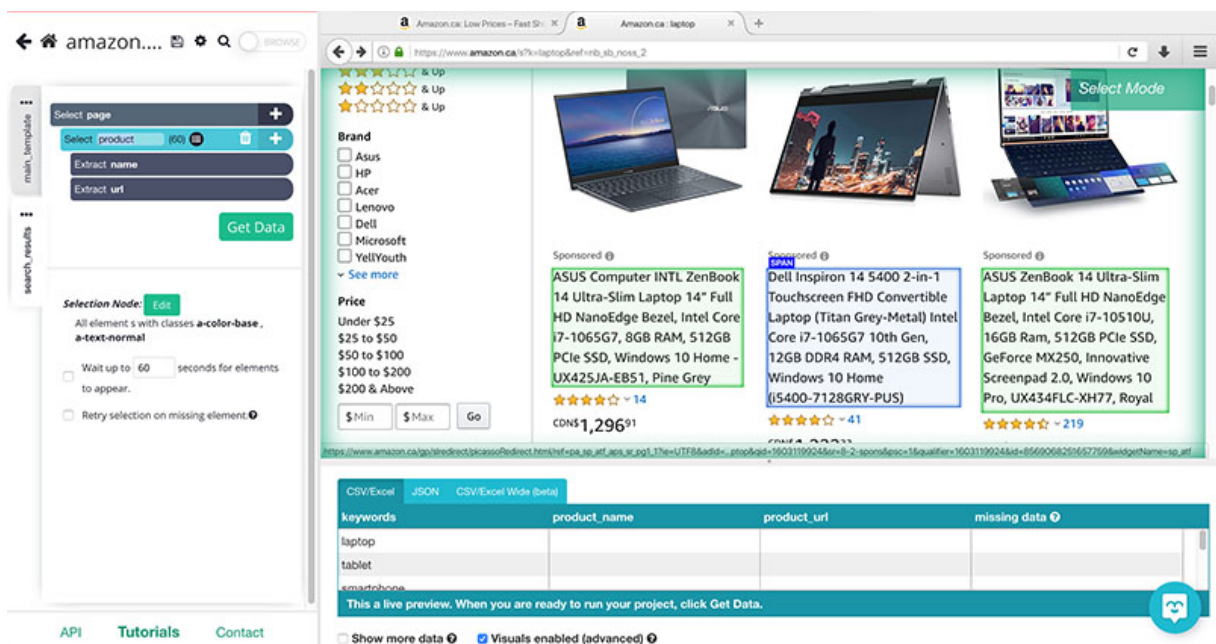
Bing, Yahoo... Ďalej umožňuje vyhľadávať kľúčové slová a z jedného kľúčového slova vytvárať kľúčové slová rôznej dĺžky. Výhodou tohto nástroja je, že nie je potrebné platiť predplatné. Stačí, ak zákazník zaplatí raz a potom môže nástroj využívať bezplatne. Nevýhodou je, že pri rozsiahlejšom prehľadávaní je nástroj veľmi pomalý. Na obrázku nižšie môžeme vidieť ukážku nástroja.

- **ParseHub** - je desktopová aplikácia, ktorá slúži na prehľadávanie webu. Dokonca sa vie popasovať aj s komplikovanými a dynamickými webovými stránkami. Príkaz na prehľadávanie webu sa vytvára v spomínanej aplikácii, no samotné prehľadávanie prebieha na serveroch ParseHub. ParseHub na prehľadávanie webu využíva technológie AJAX, JavaScript, cookies... Využíva taktiež strojové učenie, pomocou ktorého je schopná čítať, analyzovať a transformovať webové dokumenty na relevantné údaje. ParseHub môžete využiť bezplatne, kde budete je možnosť nastaviť 5 projektov prehľadávania zadarmo. Platený plán umožňuje nastaviť aspoň 20 projektov prehľadávania. ParseHub ponúka možnosť vzdelávania sa v oblasti web scrapingu pomocou interaktívnych videí a kurzov. Na obrázku nižšie môžeme vidieť malú ukážku ParseHubu.

Analýzou dostupných riešení sme zistili, že existuje niekoľko aplikácií alebo webstránok, ktoré sa venujú web scrapingu. Veľa nástrojov zameraných na webové prehľadávanie obsahuje viaceré funkcionality, ktoré nie sú až tak potrebné. Z toho dôvodu sa potom



Obr. 5: ScrapeBox



Obr. 6: ParseHub

navyšuje cena za predplatné. Taktiež niektoré z nich pôsobia veľmi zložito, čo by ľudia, ktorí nie sú technicky zdatní, nemuseli vedieť obsluhovať. Preto je našim cieľom vytvoriť jednoduchú webovú aplikáciu, s ktorou bude vedieť pracovať každý.

4.2 Analýza dát

Analýza údajov je proces kontroly, „čistenia“, transformácie a modelovania údajov s cieľom objaviť a vyťažiť len užitočné informácie z veľkého objemu dát. Precízna a kvalitná analýza údajov a tvorba datasetov je „alfa a omega“ každého úspešného projektu.

V tejto časti si opíšeme postupy pri analýze portálov a hľadania vhodných dát (článov), z ktorých by sme mohli čerpať pri tvorbe datasetu, potrebného na natrénovanie modelu.

4.2.1 EDA - Exploratory Data Analysis

Prieskumná analýza údajov (z angl. Exploratory Data Analysis) sa vzťahuje na proces vykonávania počiatočného skúmania údajov a jej cieľom je získať všeobecný pohľad na dáta s ktorými budeme pracovať.

EDA sme začali analyzovaním portálov na základe nasledovných kľúčových slov:

1. „ohlásil/a“ | „novú/nové/novou“ | „významnú/é“ | „zahraničná/ú“:

- **investícia/investíciu/investície**

2. „ohlásil/a“ | „novú/nové/novou“ | „významnú/é“:

- **projekt**
- **investičný zámer**
- **rozšírenie projektu**
- **expanzia/expanziu**
- **vstup na Slovenský trh**

3. „vybudovanie“ | „ohlásenie“ | „výstavba“:

- **výrobná linka**
- **rozšírenie výroby**
- **zvýšenie výrobných kapacít**

4. „významnú“ | „ohlásil/a“ | „prebiehajúca“ | „pripravovaná“:

- akvizícia/akvizíciu

5. „ohlásil/a“:

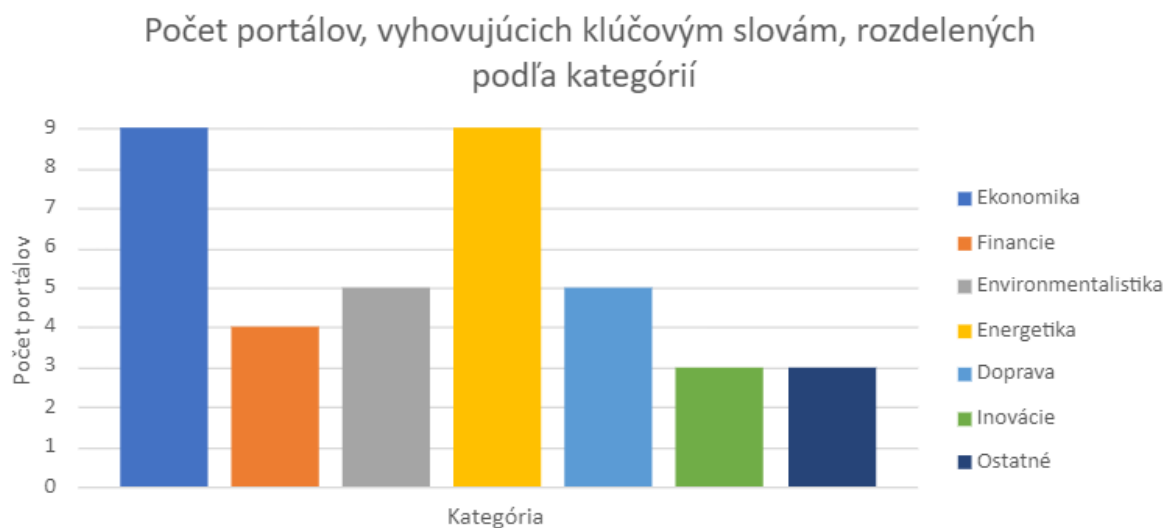
- konkurz
- reštrukturalizácia/u

6. bez sprievodných slov:

- kúpa/u spoločnosti / predaj spoločnosti
- M&A/SPV
- odkúpenie akcií/obchodných podielov
- emisia/u dlhopisov
- emisia/u cenných papierov
- prichádza firma
- výstavba nového
- obchodná príležitosť
- investičná príležitosť
- crowdfunding
- investičná akcia
- príchod investora
- investor
- plánuje investovať

Okrem kľúčových slov, sme na portáloch kontrolovali aj povolenie zobrazenia zdrojového kódu stránky, či čítanie niektorých článkov nevyžaduje prihlásenie na portál, poplatky a podobe.

V histograme je zobrazený počet nájdených portálov spĺňajúcich naše kritéria, ktoré sme na začiatok rozdelili do siedmich základných kategórií, podľa toho, na akú oblasť sa prevažne zameriavajú články, ktoré obsahujú. Ako môžeme na histograme vidieť, nepodarilo sa nám nájsť dostatočné množstvo portálov, z ktorých by sme boli schopní vytvoriť relatívne veľký dataset.



Obr. 7: Histogram zanalyzovaných portálov

Manuálne prehľadavanie internetu je zdĺhavý a neefektívny proces, pretože pri zadaní len jedného kľúčového slova sme z prehliadača dostali približne milión výsledkov - manuálna kontrola takéhoto obrovského množstva informácií, ktorým v súčasnosti disponuje internet, je nad ľudské sily.

Data mining, t.j. proces zozbieravania dát sme zoptimalizovali a to využitím knižnice **Google Search**. Viac v kapitole 6.1.

Prvotná analýza nám dala prehľad o dostupných portáloch a možnosť overiť si správnosť výstupov zo spomínanej knižnice Google Search.

5 Technológie

Keďže na trhu existuje široká škála zaujímavých technologických riešení, rozhodli sme sa niektoré najznámejšie dostupné technológie zanalyzovať a následne implementovať to najlepšie riešenie pre náš softvér.

5.1 Data Processor

Data Processor je v našom softvéri modul slúžiaci na získavanie surových dát, ich následnú analýzu a finálne spracovanie. Bude využívať technológie popísané v nasledujúcich podkapitolách.

5.1.1 Google Search

Pythonovská knižnica GoogleSearch je veľmi jednoduchá na použitie a zároveň veľmi praktická a užitočná. Funguje na princípe obyčajného prehliadača, čo vyplýva aj z jej názvu. Metóda search z tejto knižnice dostane ako parameter slovo/slová na základe ktorých prehľadáva prehliadač Google. Výhodná je taktiež možnosť nastavenia filtrov na doménu, jazyk, čas a podobne. Táto knižnica nám veľmi uľahčila prácu. Problematiku prvej analýzy sme opisovali v kapitole 4.1.

5.1.2 Fuzzy Logic

Keďže slovenčina je rôznorodý jazyk a jedno slovo môže mať veľa podôb a tvarov, rozhodli sme sa využiť metódy fuzzy logiky.

Fuzzy logika využíva matematické vzorce na ohodnocovanie podobnosti slov. Čo v praxi znamená, napríklad slová investícia/ivestíciu/investíciou vyhodnotí ako rovnaké slovo, priradením vysokého skóre.

5.1.3 Natural Language Processing

Pri spracovaní prirodzeného jazyka (NLP) ide o schopnosť porozumieť textu a hovorovým slovám v podstate rovnakým spôsobom, akým to dokážu ľudské bytosti.

NLP kombinuje modelovanie ľudského jazyka na základe pravidiel so strojovým učeníím a modelmi hlbokého učenia. Tieto technológie spoločne umožňujú počítačom spracovávať ľudský jazyk vo forme textových alebo hlasových údajov a „pochopiť“ jeho plný význam.[1]

Spracovanie prirodzeného jazyka obsahuje dve fázy:

- **Predspracovanie údajov** - zahŕňa prípravu a čistenie textových údajov, ktoré sa dajú do použiteľnej formy a zvýrazní slová v texte pre algoritmus. Na toto existuje niekoľko spôsobov, ako to urobiť. Nižšie sú popísané niektoré z nich:
 - **Tokenizácia** - text sa rozdelí na viaceré menšie časti, s ktorými potom algoritmus pracuje.
 - **Odstránenie zastavovacích slov** - v texte sa nájdu kľúčové slová, ktoré poskytujú najviac informácií o texte a zároveň sa odstránia bežne používané slová.
 - **Lemmatizácia a stemming** - zredukovanie slov v texte na koreňové tvary.
 - **Značkovanie slov** - označenie slov v texte na základe slovných druhov.
- **Vývoj algoritmu** - po predspracovaní údajov sa vytvorí algoritmus na ich spracovanie. Existuje viacero takýchto algoritmov, no medzi najhlavnejšie patria:
 - **Rules-based system** - systém založený na pravidlách využíva starostlivo navrhnuté lingvistické pravidlá. Tento typ systému sa používa od začiatku vývoja spracovania prirodzeného jazyka.
 - **Machine learning-based system** - systém založený na algoritmoch strojového učenia využíva štatistické metódy. Na základe tréningu dát, ktoré dostanú, sa učia vykonávať úlohy. Algoritmy spracovania prirodzeného jazyka v spojení so strojovým učením, hlbokým učením a neurónovými sieťami zdokonaľujú svoje pravidlá pomocou opakovaného učenia.

Medzi najhlavnejšie funkcie, ktoré vykonávajú algoritmy na spracovanie prirodzeného jazyka sú:

- **klasifikácia textu** - priradenie značiek k textom na zaradenie do rôznych kategórií.
- **extrakcia textu** - automatické zhrnutie textu a nájdenie dôležitých údajov. Napríklad extrakcia kľúčových slov, ktorá vyberie z textu najdôležitejšie slová z textu.
- **strojový preklad** - proces, kde počítač preloží text z jedného jazyka do druhého.

- **generovanie prirodzeného jazyka** - automatické vytváranie obsahu pomocou algoritmov na spracovanie prirodzeného jazyka, ktorý zanalyzuje neštruktúrované údaje.

Výhody:

- Zlepšenie komunikácie medzi človekom a počítačom.
- Schopnosť automaticky vytvárať čitateľný text z väčšieho, zložitejšieho pôvodného textu.
- Užitočné pre vývoj osobných asistentov, ako je napríklad Alexa, pretože im umožňuje porozumieť hovorenému slovu.
- Umožňuje organizáciám používať chatboty na zákaznícku podporu.

V projekte sme na predspracovanie údajov využili metódu na odstránenie zastavovacích slov. Na vývoj samotného algoritmu sme použili systém založený na algoritmoch strojového učenia. Programovací jazyk Python poskytuje širokú škálu nástrojov a knižníc na prácu s NLP, napríklad na analýzu textu v neštruktúrovaných dátových formátoch vrátane HTML, na identifikáciu kľúčových slov, kategórií a podobne.

5.1.4 Machine learning

Machine learning, po slovensky strojové učenie, je typ umelej inteligencie (AI), ktorý umožňuje softvérovým aplikáciám presnejšie predpovedať výsledky bez toho, aby boli na to explicitne programované. Algoritmy strojového učenia vytvárajú matematický model založený na vzorových údajoch, známe ako trénovacie dáta, pomocou ktorého dokážu predpovedať nové výstupné hodnoty.[2][3]

Výhody:

- **Automatizácia všetkého** - strojové učenie pomáha znížiť pracovné zaťaženie a čas, čo uľahčuje prácu ľuďom.
- **Široká škála využitia** - strojové učenie je možné aplikovať na akúkoľvek z hlavných oblastí. Či už je to oblasť medicíny, bankovníctva, obchodu, techniky alebo vedy, všade dokáže byť nápomocné.

- **Zlepšenie do budúcnosti** - keďže strojové učenie je stále nová technológia a neustále sa vyvíja, má veľkú perspektívu do budúcnosti. V budúcnosti by sa mohla stať špičkovou technológiou.
- **Efektívne spracovanie údajov** - strojové učenie disponuje mnohými faktormi, vďaka ktorým je veľmi spoľahlivé. Pomocou strojového učenia môžeme spracovať akýkoľvek typ údajov, čo niektoré bežné systémy nedokážu.

Nevýhody:

- **Možnosť vysokej chyby** - keďže algoritmy strojového učenia pracujú s obrovským množstvom údajov, často dochádza k chybám pri trénovaní a testovaní dát. Vo väčšine prípadov je úplné odstránenie chýb prakticky nemožné.
- **Výber algoritmu** - predtým ako si vyberieme jeden konkrétny algoritmus, ktorý použijeme, je potrebné na dostupných údajoch otestovať, ako sa daný algoritmus správa. Podľa presnosti výsledkov si následne vyberieme ten správny.
- **Nekonzistentnosť dát** - tým, že pri strojovom učení pracujeme s obrovským množstvom dát, je potrebné ich často aktualizovať. Ak by sme tak neurobili, môže nastať nekonzistentnosť dát medzi starými a novými údajmi. Vtedy môžeme dostať odlišné výsledky, čo nie je pre vybraný algoritmus vhodné.
- **Čas a priestor** - mnoho algoritmov strojového učenia môže zaberať viac času, ako si dokážeme predstaviť. Všetko závisí od toho, aké veľké sú údaje a čo od algoritmu očakávame. To isté platí aj pridelený dátový priestor a vyššiu spotrebu CPU.
- **Peniaze** - projekty založené na strojovom učení sú zvyčajne dosť drahé. Vyžadujú si tím odborníkov v danej oblasti a tiež vysoké nároky na softvérovú infraštruktúru.

5.2 Backend

Ako už z názvu vyplýva, backend je časť webu, ktorú užívateľ nevidí. Zodpovedá za ukladanie, organizáciu údajov a zabezpečenie toho, aby všetko na strane klienta fungovalo. Komunikuje s frontendom, odosiela a prijíma informácie.

5.2.1 Laravel

Laravel je open-source PHP, MVC framework určený na uľahčenie a zrýchlenie vývoja webových aplikácií pomocou zabudovaných funkcií. Ide o najrozšírenejší PHP framework na svete. [4].

Výhody:

- **Modularita:** ponúka 20 integrovaných modulov a knižníc, ktoré sú kombinované. Vzhľadom k tomu, že každý modul je pokrytý Composerom značne uľahčuje aktualizácie.
- **Rozšírené zabezpečenie:** poskytuje zvýšenú bezpečnosť pre dáta a používateľov. Neumožňuje prístup plateným zdrojom alebo neoprávneným používateľom.
- **Vynikajúca integrácia poštových služieb:** ponúka silné a čisté API, vďaka knižnici Swift Mailer.

Nevýhody:

- **Žiadne zabudované podporné funkcie:** na rozdiel od Django má Laravel menšiu zabudovanú podporu, pretože ide o malý framework. Tento problém je možné vyriešiť pomocnými programami.
- **Niektoré vylepšenia môžu byť problematické:** s dlhodobou podporou má vo všeobecnosti niekoľko problémov týkajúcich sa verzií. Aktualizácie môžu spôsobiť menšie problémy. To znamená, že určitý kód v novej verzii nemusí fungovať a je potrebné ho prepísať.

5.2.2 Django

Django je Web framework na vysokej úrovni v jazyku Python, ktorý podporuje rýchly vývoj a čistý, pragmatický dizajn. Ide o open-source MVC framework. Spracované podľa [5].

Výhody:

- **Bezpečnosť:** pomáha vývojárom vyhnúť sa bežným problémom so zabezpečením proti útokom ako je napríklad clickjacking, skriptovanie medzi servermi, SQL injection, CSRF.

- **Algoritmy strojového učenia:** je kompatibilný s niektorými výkonnými knižnicami strojového učenia ako napríklad Pytorch a Numpy. Výpočtové a štatistické schopnosti Django z neho robia ideálnu platformu pre aplikácie strojového učenia.
- **Batteries included:** namiesto písania kódu môžu vývojári použiť pripravené balíčky. Vďaka nim môžete implementovať autentifikáciu, prostredie pre administrátora, testovanie a podobne.

Nevýhody:

- **Chýbajúce konvencie:** Django nemá súbor zásad, ktoré môžu vývojári dodržiavať.
- **Menej vhodný na malé projekty:** majú len niekoľko funkcií a požiadaviek. Django sa častejšie používa na projekty, ktoré je potrebné škálovať a kde je možné využiť plný potenciál rámca s jedinečnými funkciami na rýchly vývoj.

5.2.3 SpringBoot

Spring Boot je open-source framework založený na jazyku Java, ktorý sa používa na vytvorenie mikro služieb. Je vyvinutý spoločnosťou Pivotal Team [6].

Výhody:

- **Uľahčuje praktické využitie známych technológií:** ORM framework, logovacie frameworky, JEE, JDK.
- **Jednoduché Unit testovanie:** tento typ testovania má mnoho výhod. Nie je nutné testovať celú aplikáciu ale iba určité časti. Umožňuje zistenie chýb, v skorších fázach vývoja, čo znižuje náklady na ich odstránenie.
- **Menej implementačného kódu:** vyhýba sa zložitej konfigurácii xml vďaka Java metódam. Má server Tomcat, ktorý pri spustení projektu vytvorí server, ktorý nastaví prístupný pre prehliadače.
- **Inverzia kontroly:** pomáha vývojárom zbaviť sa nutnosti písania samostatnej kompilačnej jednotky alebo samostatného zavádzača tried na zvládnutie výnimiek. Ak chceme pridať svoje vlastné správanie, musíme rozšíriť triedy frameworku alebo vytvoriť nové.

Nevýhody:

- **Veľkosť súboru nasadenia:** inštaluje mnoho ďalších závislostí (ktoré sa často nepoužívajú), čo zvyšuje veľkosť súboru nasadenia.

- **Aktualizácia:** môže byť dosť náročné aktualizovať váš starší kód. Tento problém môžete prekonať pomocou nástrojov, ako je Spring Boot CLI, ktoré vám pomôžu previesť váš starší kód.

5.2.4 Vyhodnotenie

Po zvážení všetkých výhod a nevýhod sme vybrali pre náš projekt SpringBoot framework postavený na Jave. Java je považovaná za najstabilnejší programovací jazyk. Má vynikajúcu reputáciu vďaka svojej efektívnej kompatibilitate medzi platformami. Navyše, Java je o niečo rýchlejšia a výkonnejšia ako PHP, na ktorom je postavený framework Laravel.

5.3 Frontend

Frontend webových stránok je všetko, s čím používateľ interaguje a čo používateľ vizuálne vidí ako prvé, keď spustí webovú aplikáciu. Podkapitola spracovaná podľa [7].

5.3.1 Angular

Angular je vyvinutý spoločnosťou Google, bol prvýkrát uvedený na trh v roku 2010. Je to rámec založený na Typescripte. Umožňuje rýchle vytváranie funkcií pomocou jednoduchých deklaratívnych šablón. Pomocou dvojsmerného viazania údajov automaticky synchronizuje údaje z používateľského rozhrania s modelom.

5.3.2 React

React je javascriptová knižnica, vyvinutá spoločnosťou Meta. React je postavený na myšlienke skladateľných komponentov, ktoré sú izolované. Veľmi uľahčuje tímovú spoluprácu, pretože rôzni ľudia môžu vytvárať a meniť rôzne komponenty, ktoré spolu dobre fungujú. V aplikácií sa komponenty nevykresľujú priamo do HTML DOM elementu. Metóda renderovania vracia popis toho, čo je potrebné vykresliť. React má rýchly a inteligentný spôsob, ako to aplikovať na DOM. Po zmene stavu v aplikácii react efektívne aktualizuje a vykresľuje len tie správne komponenty.

5.3.3 Vue

Vue je frontendový framework, tiež známy ako Vue.js. Za posledné tri roky zaznamenal výrazný posun v popularite, aj keď zatiaľ nemá veľkú komunitu. Má najmenšiu veľkosť a je ľahko pochopiteľný čo môže byť hlavný dôvod popularity. Podobne ako React,

je založený na komponentoch.

5.3.4 Vyhodnotenie

React a jeho rozdelenie funkcií do jednotlivých komponentov, ktoré promptne reagujú na zmenu stavov, popularita vo svete, prehľadná syntax, dobré vyhliadky do budúcnosti nás presvedčili o jeho výbere pre našu aplikáciu.

5.4 Databáza

V tejto sekcii si opíšeme základné informácie o databázach vhodných pre naše riešenie, ich výhody a nevýhody. Implementácia databázy je nutnou podmienkou pre náš softvér. Kapitola je spracovaná podľa [8], [9], [10], [11] a [12].

5.4.1 SQL vs NoSQL

SQL databázy používajú štruktúrovaný dotazovací jazyk alebo SQL a majú preddefinovanú schému na definovanie a manipuláciu s údajmi. SQL je jedným z najuniverzálnejších a najrozšírenejších dostupných dopytovacích jazykov, čo z neho robí bezpečnú voľbu pre mnohé prípady použitia.

Je ideálny pre zložité otázky. SQL však môže byť niekedy príliš obmedzujúci - na určenie štruktúry údajov musíte použiť preddefinované schémy, aby ste s nimi mohli pracovať. Všetky vaše údaje musia mať rovnakú štruktúru. Tento proces si vyžaduje značnú predbežnú prípravu. Ak by ste niekedy chceli zmeniť štruktúru údajov, bolo by to ťažké a narušilo by to celý Váš systém.

NoSQL databázy majú dynamické schémy pre neštruktúrované dáta a dáta sú ukladané mnohými spôsobmi. Pre svoje údaje môžete použiť stĺpcovo orientovaný, dokumentový, grafový alebo kľúčový ukladač priestor. Táto flexibilita znamená, že:

- dokumenty môžete vytvárať bez toho, aby ste museli najprv definovať ich štruktúru,
- každý dokument môže mať svoju jedinečnú štruktúru,
- syntax sa môže líšiť od databázy k databáze,
- polia môžete pridávať priebežne.

SQL databázy sú založené na tabuľkách, zatiaľ čo databázy NoSQL ukladajú dáta ako dokumenty, key-value páry, grafy alebo široké stĺpce.

5.4.2 MySQL

Výhody:

- **Ideálne pre konzistentné dátové systémy:** s RDBMS na báze SQL vaše informácie zostanú v štruktúre, ktorú ste pôvodne vytvorili. Ak nepotrebuje dynamický informačný systém pre veľké množstvo údajov - a nemáte dočinenia s mnohými typmi údajov - RDBMS ponúka veľkú rýchlosť a stabilitu.
- **Lepšie možnosti podpory:** databázy RDBMS existujú už viac ako 40 rokov, je jednoduchšie získať podporu, prídavné produkty a integrovať údaje z iných systémov.

Nevýhody:

- **Výzvy v oblasti škálovateľnosti a ťažkostí s shardovaním:** RDBMS sa v porovnaní s databázami NoSQL ťažšie škálujú v reakcii na masívny rast. Tieto databázy tiež predstavujú výzvy, pokiaľ ide o sharding. Sharding je proces rozdelenia veľkej databázy na menšie časti pre jednoduchšiu správu. Ak máte dočinenia s konzervatívnou databázou, od ktorej neočakávate, že sa v nadchádzajúcich rokoch veľa zmení, výzvy týkajúce sa shardingu a škálovania súvisiace s riešením RDBMS sa vás nemusia týkať. Na druhej strane, ak plánujete škálovanie a rast v nasledujúcich rokoch, nerelačný databázový systém (založený na NoSQL) by mohol byť pre vaše potreby vhodnejší.
- **Menej efektívny s formátmi NoSQL:** väčšina RDBMS je teraz kompatibilná s formátmi údajov NoSQL, ale nepracujú s nimi tak efektívne, ako nerelačné databázy.

5.4.3 MongoDB

Výhody:

- **Vynikajúce na spracovanie analýzy "big data":** hlavným dôvodom, prečo sú databázy NoSQL čoraz obľúbenejšie, je to, že odstraňujú prekážku potreby kategorizovať a aplikovať prísne štruktúry na obrovské množstvo informácií.
- **Žiadne obmedzenia typov údajov, ktoré môžete ukladať:** databázy NoSQL vám dávajú neobmedzenú slobodu ukladať rôzne typy údajov na rovnaké miesto. To ponúka flexibilitu pri pridávaní nových a rôznych typov údajov do vašej databázy kedykoľvek.

- **Jednoduchšie škálovanie:** databázy NoSQL sa ľahšie škálujú. Sú navrhnuté tak, aby ich bolo možné bez väčších problémov rozdeliť do viacerých dátových centier.
- **Nevyžaduje sa žiadna príprava údajov:** keď nie je čas na návrh zložitého modelu a potrebujete spustiť databázu rýchlo, nerelačné databázy ušetria veľa času.

Nevýhody:

- **Ťažšie nájsť podporu:** komunita NoSQL nemá za sebou 40 rokov histórie a vývoja, môže byť ťažšie nájsť skúsených používateľov, keď potrebujete riešiť problémy.
- **Nedostatok nástrojov:** keďže systém je relatívne nový v porovnaní s riešeniami RDBMS založenými na SQL, nie je k dispozícii toľko nástrojov na pomoc pri testovaní výkonu a analýze.
- **Výzvy v oblasti kompatibility a štandardizácie:** novším databázovým systémom NoSQL tiež chýba vysoký stupeň kompatibility a štandardizácie, ktorý ponúkajú alternatívy založené na SQL. Možno zistíte, že údaje vo vašom systéme správy nerelačných databáz sa nedajú ľahko integrovať s inými produktmi a službami.

5.4.4 Elasticsearch

Elasticsearch je NoSQL databáza, čo znamená, že ukladá údaje neštruktúrovaným spôsobom a teda na ich dotazovanie nemôžeme použiť SQL.

Elasticsearch je vhodné použiť pri probléme textového vyhľadávania, kde tradičné databázy RDBMS nefungujú veľmi dobre. Je vysoko prispôsobiteľný a rozšíriteľný pomocou doplnkov. Taktiež je to vhodná DB pri reportingu a vizualizácii dát, pretože s Elasticsearch možno použiť API Kibana, ktoré je ideálne na:

- vyhľadávanie, prezeranie a vizualizácia údajov uložených v Elasticsearch,
- analýzu údajov prostredníctvom vytvárania stĺpcových grafov, koláčových grafov, tabuliek, histogramov a máp, ktoré sa potom zdieľajú prostredníctvom prehliadača, aby poskytovali analytické zobrazenia v reálnom čase.

Nevýhodou je, že Elasticsearch nie je dobrým dátovým úložiskom ako je napr. MongoDB. Funguje dobre len pre malé prípady použitia.

5.4.5 Vyhodnotenie

Po zhodnotení všetkých výhod a nevýhod analyzovaných databáz sme sa rozhodli použiť pre náš softvér dve databázy, konkrétne: Elasticsearch a MySQL.

Z vyššie uvedených dôvodov spĺňa Elasticsearch všetky požiadavky potrebné na vizualizáciu údajov z reportov.

MySQL databázu využijeme na ukladanie registrovaných užívateľov v systéme, pretože ponúka rýchlosť, stabilitu a je postačujúca pri menších množstvách dát ktoré vyžadujú pevnú štruktúru.

6 Implementácia

6.1 Architektúra

6.1.1 REST

REST je architektonický štýl, ktorý definuje súbor obmedzení, ktoré sa majú použiť na vytváranie webových služieb. Technológia REST je vo všeobecnosti veľmi populárna a obľúbená, pretože je jednoduchá a flexibilná.

Charakteristika REST architektúry podľa [13]:

1. **Separácia klienta a servera:** kód na strane klienta možno kedykoľvek modifikovať bez ovplyvnenia prevádzky servera a naopak. Zlepšuje sa tak flexibilita rozhrania naprieč platformami a navyše je každej zložke umožnený nezávislý vývoj.
2. **Komunikácia medzi klientom a serverom:** klienti posielajú požiadavky (z angl. requests) na získanie alebo úpravu zdrojov (z angl. resources) a servery posielajú odpovede (z angl. responses) na tieto požiadavky.
3. **Request:** REST vyžaduje, aby klient zadal požiadavku na server s cieľom získať alebo upraviť údaje na serveri. Request vo všeobecnosti obsahuje:
 - (a) HTTP sloveso, ktoré definuje, aký druh operácie sa má vykonať:
 - i. GET - získanie zdroja,
 - ii. PUT - vytvorenie nového zdroja,
 - iii. POST - úprava zdroja,
 - iv. DELETE - odstránenie zdroja.
 - (b) hlavičku, ktorá umožňuje klientovi odovzdať informácie o požiadavke. V hlavičke požiadavky klient odošle typ obsahu, ktorý je schopný prijať zo servera. Zapezpečí sa tým, že server neposiela údaje, ktoré klient nemôže pochopiť alebo spracovať.
 - (c) cestu k zdroju, na ktorom by sa mala operácia vykonať. Cesty by mali byť navrhnuté tak, aby boli čitateľné, ľahko pochopiteľné a pomohli klientovi rozpoznať, čo sa aktuálne deje.
 - (d) nepovinné telo správy obsahujúce údaje.

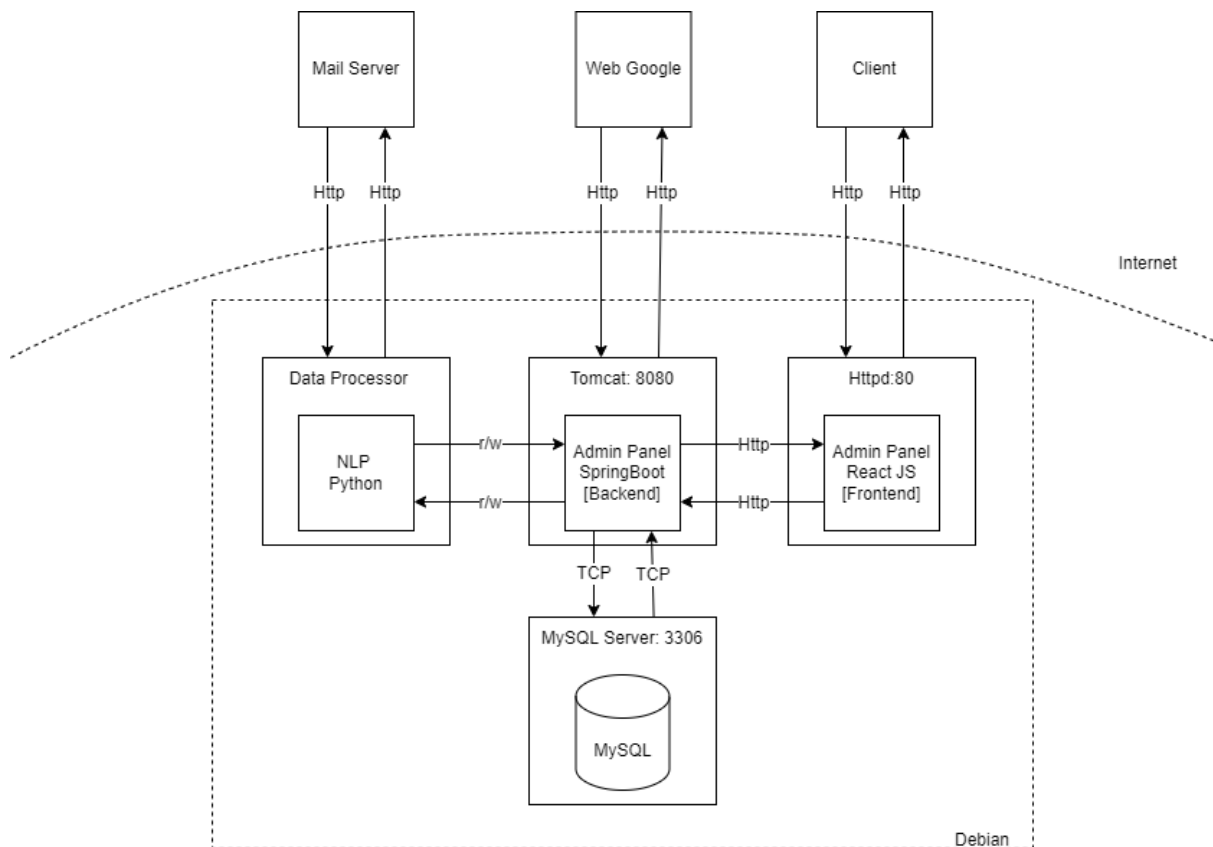
4. **Response:** odpoveď na požiadavku. Typ obsahu, ktorý server odošle späť v odpovedi, by mal byť jednou z možností, ktoré klient zadal v požiadavke. V našej aplikácii používame anotáciu RestController, ktorá automaticky odpoveď konvertuje do JSON formátu. Odpovede zo servera obsahujú stavové kódy, ktoré upozorňujú klienta na informácie o úspechu príp. neúspechu operácie. Najbežnejšie používané sú:

Kód	Stav	Popis
200	OK	Štandardná odpoveď na úspešnú HTTP požiadavku.
201	CREATED	Štandardná odpoveď na HTTP požiadavku, ktorá viedla k úspešnému vytvoreniu položky.
204	NO CONTENT	Štandardná odpoveď pre úspešné HTTP požiadavky, kde sa v tele odpovede nič nevracia.
400	BAD REQUEST	Požiadavku nemožno spracovať z dôvodu nesprávnej syntaxe požiadavky, nadmernej veľkosti alebo inej chyby klienta.
403	FORBIDDEN	Klient nemá povolenie na prístup k danému zdroju.
404	NOT FOUND	Zdroj sa momentálne nenašiel. Je možné, že bol vymazaný, alebo ešte neexistuje.
500	INTERNAL SERVER ERROR	Všeobecná odpoveď na neočakávané zlyhanie, ak nie sú k dispozícii žiadne konkrétnejšie informácie.

Tabuľka 1: Response stavové kódy

6.1.2 Model implementovanej aplikácie

Diagram na obr. 8 je „kostra“ implementovaného systému a znázorňuje interakcie medzi jednotlivými komponentmi aplikácie, servermi, používateľským rozhraním a databázou.



Obr. 8: Model aplikácie

6.2 Doménová časť

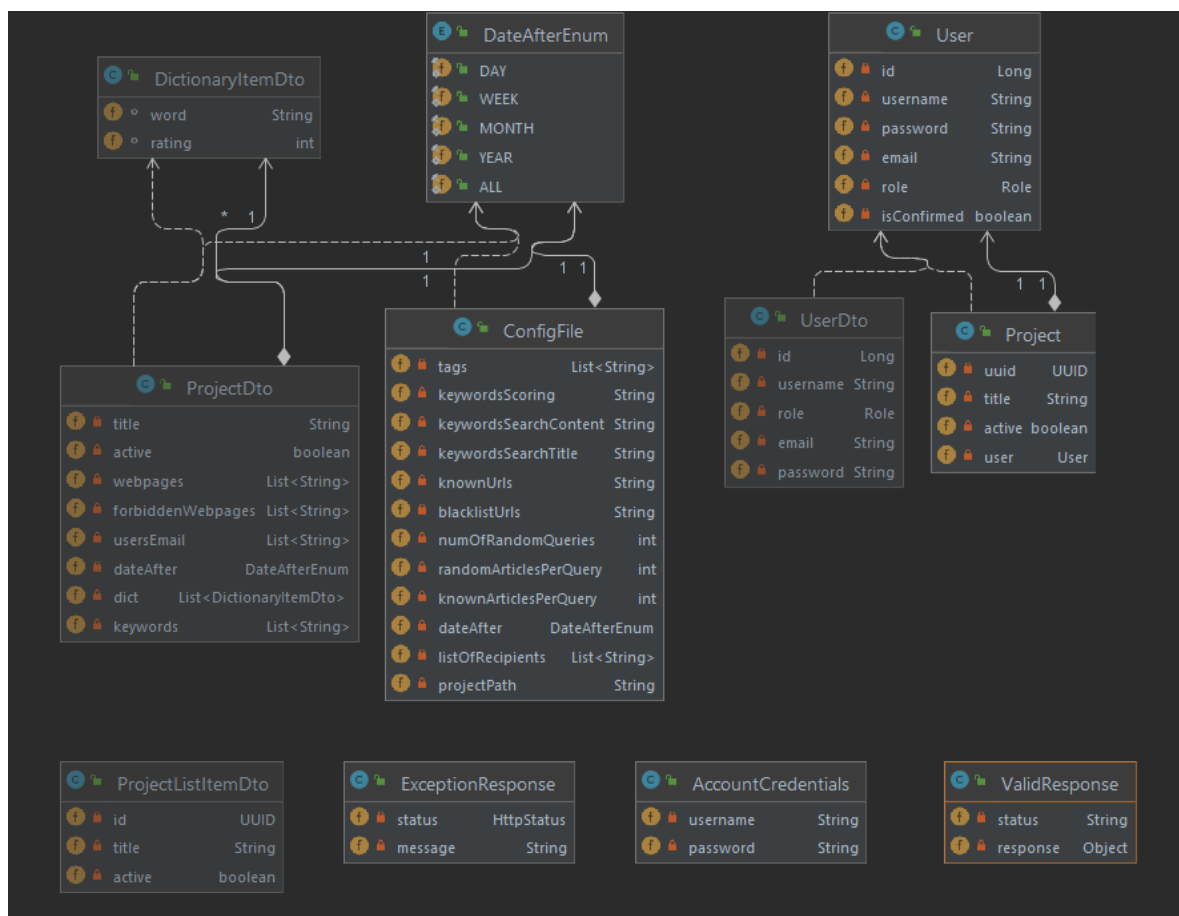
V našej aplikácii sme sa rozhodli využívať JPA špecifikáciu, ktorá definuje ako uchovávať údaje v Java aplikáciách. Ako implementáciu JPA sme zvolili Hibernate framework. Podľa informácií zo zdroja [14] ide o najobľúbenejší ORM framework.

Persistence nám zaisťuje stálosť uložených údajov a zabezpečuje to, aby vytvorený objekt v Jave mal rovnakú štruktúru v databáze. Kvôli tomuto JPA špecifikuje objektovo-relačné mapovanie, ktoré presne určuje, akým spôsobom budeme mapovať entity z databázy na objekty v Jave. Naš zvolený Hibernate framework umožňuje výkonné objektovo-relačné mapovanie. Okrem toho na pozadí sám vytvára sql príkazy nad databázou. To znamená, že pri hľadaní tabuľky podľa id, ukladaní záznamu alebo pri vylistovaní všetkých

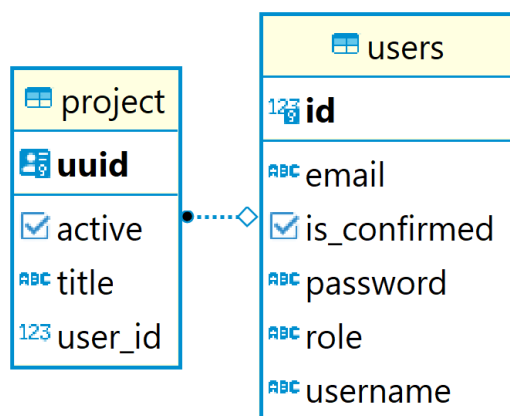
záznamov nemusíme písať žiadny sql kód, ale hibernate nám poskytuje rôzne zabudované funkcie a metódy.

V našej aplikácii sme vytvorili všetky potrebné POJO objekty a tie sme označili anotáciou `Entity` z balíčka `javax.persistence`. Z tohto balíčka používame aj ďalšie anotácie, ktoré nám umožňujú ORM mapovanie. Primárny kľúč označujeme anotáciou `Id` a jeho hodnota je autogenerovaná. Na definíciu vzťahov medzi entitnou triedou `Project` a `User` používame anotáciu `ManyToOne`. Cez anotáciu `Column` a `Table` vieme zmeniť názov stĺpca a entity. Na prácu s údajmi, používame CRUD repozitáre. Tieto repozitáre poskytujú základné funkcie. Zložitejšie dopyty na databázu vytvárame s použitím anotácie `Query` s využitím jazyka JPQL, ktorý je veľmi podobný jazyku SQL. Hlavný rozdiel je v tom, že SQL pracuje s relačnými tabuľkami a JPQL s java objektami.

Diagram tried na obrázku č.9, sme vygenerovali v našom používanom vývojovom prostredí IntelliJ IDEA. Entitno relačný diagram na obr.č. 10 nám vygeneroval nástroj na správu databáz DBeaver.



Obr. 9: Diagram tried



Obr. 10: Entitno-relačný diagram

6.3 Autentifikácia

Jednou z funkcionalít nášho systému je registrácia a prihlásenie. Na bezpečne posielanie requestov sme využili JWT token a Spring Security framework.

Ide o krátky zakódovaný text, ktorý sa posiela v hlavičke požiadavky medzi klientom a serverom. Token sa skladá z troch častí oddelených bodkami. Hlavička tokenu definuje typ tokenu a hash algoritmus pod ktorým je uložené heslo. V druhej časti sú prihlasovacie údaje a podpis sa používa na zabezpečenie integrity tokenu. To znamená jeho nemennosť.

```
eyJhbGciOiJIUzUxMiJ9.eyJzdWIiOiJ0ZXN0IiwiaXNwIjoxNjUyMTk2NjQ4LCJyb2xlIjoiaUk9MRV9VU0VSIn0.T79nrfD_-DS1BDk1R1QX93cuQIh0SNH4w90ptEBhaDzNCIesM3T9FGgKXaW4Txs1CYWpGyoyjkVie08LVqyqzg|
```



Obr. 11: Ukážka dekódovaného tokenu z našej aplikácie (www.jwt.io)

Pri prihlásení do aplikácie sa odošle požiadavka metódou POST na URI api/login.

V tele požiadavky sa odošle prihlasovacie meno a heslo. Systém tieto prihlasovacie údaje overí a ak sú korektné odošle klientovi platný JWT token. Pri každej požiadavke, kde je nutné byť autentifikovaný klient odosiela JWT, ktorý systém zakaždým kontroluje. Ak je JWT validný a s potrebnými oprávneniami systém sprístupní údaje o ktoré klient žiada, v opačnom prípade prístup zamietne. Tým docielime, že každý používateľ uvidí iba svoje projekty a k ostatným nebude mať prístup. Vďaka pripraveným rolám vieme jednoducho rozširovať funkcionalitu aplikácie.

6.4 Práca so súbormi a s databázou

Pri vytvorení projektu sa v databáze uloží nový záznam v tabuľke project a vytvorí sa relácia na prihláseného používateľa (tabuľka user). Používateľ zadáva emaily používateľov, na ktoré sa budú odosielať výsledné reporty, webstránky z ktorých sa budú sťahovať údaje a zakázané webstránky, ktoré budú vynechané z prehľadávania a taktiež požadované kľúčové slová. Tieto údaje prichádzajú na backend ako pole stringov, ktoré sa následne uloží do textového súboru vo formáte 1 riadok = 1 string. Používateľ okrem iného zadáva aj slovník s ohodnotenými slovami. Tieto ohodnotené slová sa uložia vo formáte csv kde sa v stĺpci words nachádza dané slovo a v stĺpci rating jeho ohodnotenie. Posledným súborom je konfiguračný json súbor. Jeho ukážka je na obr. 13. Prácu s json súborom nám umožňuje trieda ConfigFile na obr. č.9. Mapovanie triedy na json a opačne, zabezpečujeme pomocou knižnice Jackson Databind, ktorá poskytuje jednoduchú prácu s takýmto objektom. Každému projektu je teda priradené unikátne UUID a každý projekt má vlastné konfiguračné súbory. Súborovú štruktúru môžeme vidieť na obr. č.12. Následne sú súbory použité ako vstupné parametre pre ďalšie spracovanie dát, ktoré je bližšie popísané v časti 7.1.6.

Úprava projektu funguje analogicky, len prepisujeme existujúce súbory novými údajmi. Poslednou funkcionalitou je zmazanie samotného priečinku s UUID projektu a taktiež zmazanie všetkých súborov, ktoré obsahuje. Okrem súborov vymažeme aj prislúchajúci údaj v databáze.

6.5 Inštancie projektov a ich management

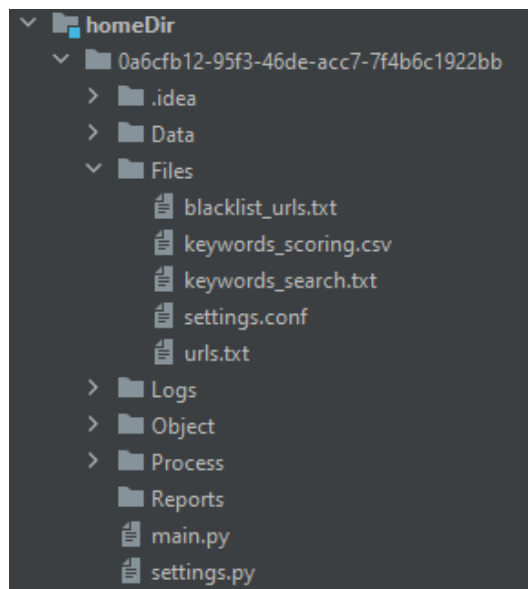
Pri práci s projektami využívame databázu pre ukladanie základných informácií o projekte, súborový systém OS a cron. Proces vytvorenia projektu začína na java endpointe, kde zaznamenáme platné údaje a vytvoríme záznam v databáze, ako sme opísali v kapitole 6.4. Následne, keď už máme UUID, vytvoríme domovskú zložku pre projekt, za pomoci unix príkazu cp. V tomto kroku sa prekopírujú súbory z pôvodnej šablóny python projektu

a prepíšu sa konfiguračné súbory, a to tak, ako je opísané v kapitole 6.4.

Následne, po vytvorení domovského adresára python inštalácie sa vytvorí cronjob, ak je projekt uložený ako aktívny (v aktuálnej implementácii je projekt vždy po vytvorení neaktívny). Cronjob má na starosti automatické púšťanie inštalácie nášho python projektu. Spúšťací čas cronjobu sa nastaví podľa užívateľských nastavení.

Pri editácii projektu je možné aktivovať/deaktivovať daný projekt. Pri danej operácii sa cronjob pridá alebo zmaže.

V prípade vyvolania operácie pre nenávratné odstránenie projektu zo systému, sa najprv zmaže cronjob (ak existoval) a následne sa zmaže adresár pre daný projekt.



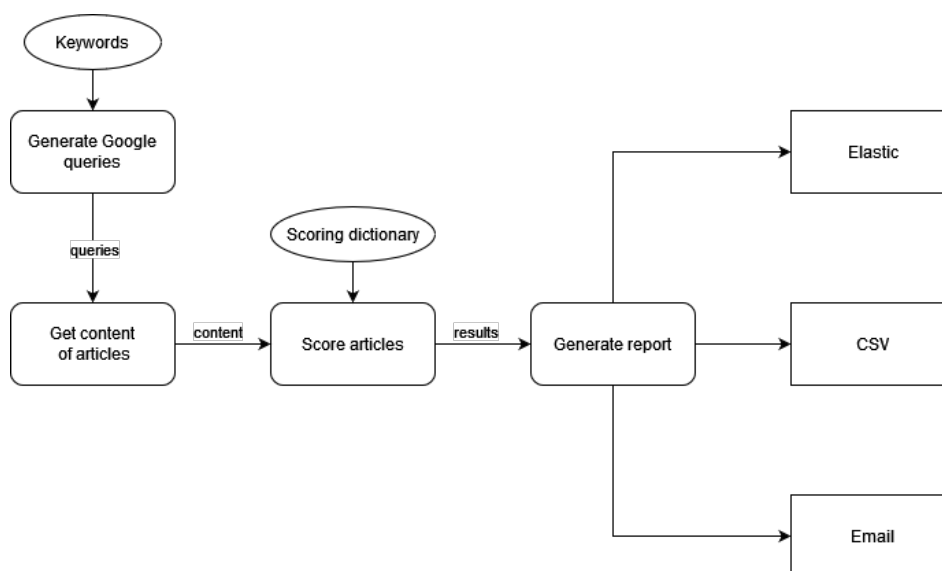
Obr. 12: Súborová štruktúra

```
{
  "tags" : [ "p", "h1" ],
  "keywords_scoring" : "7a48e58d-753b-429a-bd6d-af2f15719018/keywords_scoring.csv",
  "keywords_search_content" : "7a48e58d-753b-429a-bd6d-af2f15719018/keywords_scoring.csv",
  "keywords_search_title" : "7a48e58d-753b-429a-bd6d-af2f15719018/keywords_scoring.csv",
  "known_urls" : "7a48e58d-753b-429a-bd6d-af2f15719018/urls.txt",
  "blacklist_urls" : "7a48e58d-753b-429a-bd6d-af2f15719018/blacklist_urls.txt",
  "num_of_random_queries" : 10,
  "random_articles_per_query" : 10,
  "known_articles_per_query" : 3,
  "date_after" : "MONTH",
  "list_of_recipients" : [ "sdfs@adasdsa.sk" ],
  "project_path" : "7a48e58d-753b-429a-bd6d-af2f15719018"
}
```

Obr. 13: Konfiguračný súbor

6.6 Python crawler

V tejto časti sa venujeme samotnému prehľadávaniu internetu podľa zadaných kľúčových slov, skórovaniu článkov podľa obsahu za pomoci zadaných slovníkov a taktiež vytváraniu reportu pre ďalšiu analýzu a spracovanie nálezov.



Obr. 14: Graf procesu zbierania a hodnotenia článkov z Google

Obrázok číslo 14 zachytáva kľúčové časti python komponentu v našom projekte. Proces začína tvorbou *Google queries*, ktoré sú vytvorené pomocou niektorých prvkov nastavielných v súbore *settings.conf*. Ide napríklad o použité kľúčové slová pri tvorbe queries, dátum alebo známe stránky, na ktorých chceme články skontrolovať. S týmito parametrami sa v programe vygeneruje požadovaný počet queries, ktoré chceme aplikovať. Program vytvára dva druhy takýchto vyhľadávacích výrazov, jeden je pre vyhľadávanie v texte článku za použitia parametra *intext* a druhý slúži na vyhľadávanie v nadpisoch pomocou parametra *intitle*.

Po získaní vyhľadávacích reťazcov môže program ďalej pokračovať a to získaním obsahu z nájdených článkov. V tomto procese výrazne pomáha knižnica *BeautifulSoup*, s ktoru vieme zo získaného HTML obsahu extrahovať len požadované časti stránky (vybrané HTML tagy) definované v *settings.conf*.

Predposledný z hlavných krokov procesu je ohodnotiť obsah článkov. Pri tomto bode sa zameriavame na skutočnosť, že články by mali patriť do požadovaného vyhľadávaného okruhu. Teda ak hľadáme články týkajúce sa investičných príležitostí, pre získanie čo najlepších výsledkov, by mal tomu zodpovedať aj slovník pre hodnotenie článku (ukážku

malej časti slovníka možno vidieť v tabuľke číslo 2). Každý článok má po dokončení hodnotenia pridelené svoje skóre, čím vyššie skóre je pridelené, tým výraznejšie sa daný článok priblížil vyhľadávanej téme a je vhodnejší pre ďalšiu analýzu. Články sú skórované na intervale $[0, \infty)$.

Keyword	Weight
investovať	100
nová na trhu	100
výstavba novej	90
príležitosť	82
obchodné	80
akcie	77
výroba	56
...	...

Tabuľka 2: Príklad skórovacieho slovníka

Na záver procesu sa výsledky lokálne zapíšu v podobe CSV súboru, ktorý je ďalej distribuovaný pomocou e-mailovej notifikácie ako príloha. Prijemcov týchto notifikácií je možné ľubovoľne nastavovať v súbore *settings.conf*. Program taktiež ponúka možnosť dáta zapísať do Elasticu.

6.7 Frontend

Na základe predošlej analýzy dostupných frameworkov bola pre budovanie používateľského prostredia na frontende použitá open-source knižnica React JS. Táto knižnica bola použitá ako základ pri vývoji jednotlivých komponentov.

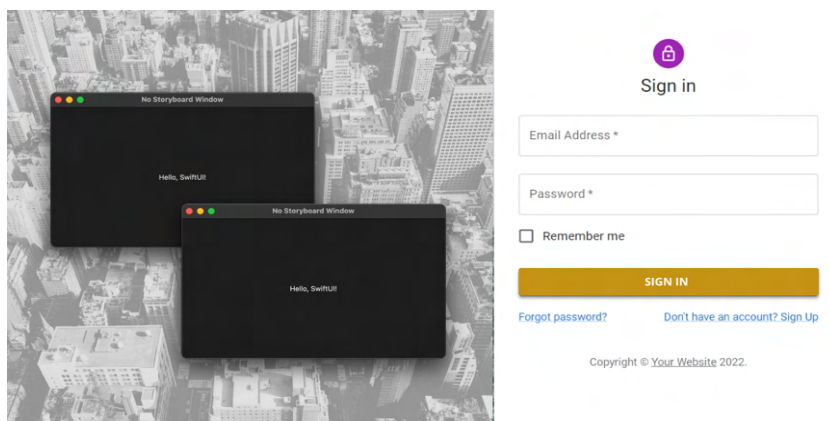
Ďalšie dodatočné knižnice pre vývoj komponentov sme získali zo softvérového registra pomocou nástroja npm. Tento nástroj slúži ako správca jednotlivých balíčkov kódu. Pomocou npm sme doinštalovali taktiež MUI, knižnicu pre efektívnejší vývoj komponentov. MUI ponúka komplexný balík nástrojov používateľského rozhrania, ktorý našu prácu zrýchli a zefektívni. V MUI máme možnosť pracovať s množstvom predpripravených komponentov a intuitívne doimplementovať vlastné riešenia a vzhľad.

Pri vývoji jednotlivých podstránok a komponentov sme sa držali vopred priprave-

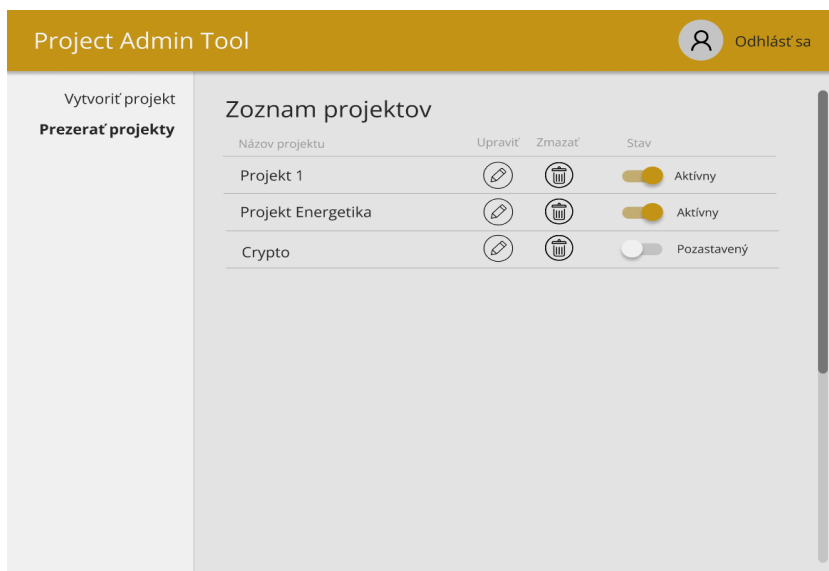
ného návrhu frontendu v nástroji Figma. Predpripravený prototyp aplikácie bol dôležitou súčasťou vývoja používateľského prostredia, aby boli dodržané všetky predpoklady UX pre intuitívne a ľahké ovládanie aplikácie.

6.7.1 UI mock-up prototypy

Pre lepšiu predstavu navrhnutého dizajnu aplikácie sme vytvorili prototypy, ktoré znázorňujú predbežný vzhľad aplikácie a rozloženie jednotlivých komponentov. Mock-upy jednotlivých stránok sú zobrazené na obrázkoch nižšie.




Obr. 15: Registrácia/Prihlásenie



Obr. 16: Zoznam projektov

Project Admin Tool

 Odhlásiť sa

Vytvoriť projekt
Prezerat' projekty

Projekt 1

1. Pridať webstránky

☐ Pridať predvolené stránky

Description

Pridať

https://dennikn.sk/slovensko/

x

https://ekonomika.sme.sk/?ref=menu

x

https://ekonomika.pravda.sk/

x

2. Pridať slovník + count

☐ Pridať predvolený slovník

Description

Pridať

Description

Pridať

investícia

100

x

nový projekt

95

x

výstavba továrne

80

x

3. Pridať používateľov

Description

Pridať

lubka@lubka.com

x

hfsdg@gamil.com

x

eee@fff.com

x

4. Ďalšie parametre

Generovať reporty do:

Date

Frekvencia generovania

☐ Každý týždeň

☐ Každý druhý týždeň

☐ Každý mesiac

Vytvoriť projekt

Obr. 17: Vytvorenie nového projektu

7 Vyhodnotenie výsledkov

Po implementácii a nasadení webovej aplikácie sme vytvorený softvér otestovali. Vygenerovali sme report, ktorý vychádzal z kľúčových slov defaultného slovníka, ktoré sme získali manuálnou analýzou portálov v prvej časti projektu. Výstup obsahoval články primárne z oblasti energetiky, dopravy, životného prostredia, automobilového priemyslu a výstavby rôznych nehnuteľností.

Všetky články z reportu sme manuálne zanalyzovali, či článok spĺňa požadované kritéria. Zistili sme, že články, ktoré mali skóre nižšie ako 600 nespĺňali naše požiadavky, pretože nemali s danou problematikou takmer nič spoločné. Naopak články, ktorých skóre bolo vyššie ako 600 sa obsahovo približovali k problematike, ktorú sme hľadali.

Najlepšie obstál článok zo stránky www.euractiv.sk (odkaz na článok [tu](#)). Ďalším v poradí bol článok týkajúci sa výstavy novej bytovky zo stránky www.teraz.sk (odkaz na článok [tu](#)). Článok získal skóre 833.

Automatizované prehľadávanie internetu je určite efektívnejšie a to hlavne z dôvodu, že je pre užívateľa časovo nenáročné a pohodlné, v porovnaní s manuálnym hľadaním dôveryhodných článkov v prehliadači. Vytvorený softvér hodnotíme kladne, pretože spĺňa všetky preddefinované požiadavky, poskytuje jednoduché, prehľadné IU, výrazne uľahčí prácu a ušetrí čas cieľovej skupine používateľov.

Jedným z možných vylepšení do budúcnosti je určite pridanie kategorizácie. Jednotlivé články vo výstupnom reporte by mohli mať špecifikovaný ďalší parameter, kategóriu, ktorej sa daný článok týka. Pre príjemcov reportu by to bolo ešte pohodlnejšie, pretože by sa vedeli zamerať na problematiku, ktorá ich zaujíma najviac.

Záver

Zimný semester:

Cieľom projektu na zimný semester bolo hlavne ujasnenie si požiadaviek so stranou VÚB prostredníctvom Ing. Michala Petráša a hlbšie ponorenie sa do problematiky. Venovali sme sa analýzam rôzneho charakteru, ujasnili sme si technológie, ktoré budeme používať a pripravili sme podklady potrebné na implementáciu v letnom semestri. Úlohy na jednotlivé týždne, ciele a celkové rozdelenie prác je zaznamenané v zápisniciach, ktoré sú priložené v tomto dokumente. Všetky stanovené ciele sme za tento semester naplnili.

Letný semester:

V letnom semestri sme sa venovali implementácii softvéru, na vyhľadávanie investičných príležitostí, čo bolo primárnym cieľom projektu. Táto časť slúži na získavanie surových dát z internetu, ich následnú analýzu, finálne spracovanie a ohodnotenie podľa nastavených kritérií. Vďaka priaznivým časovým podmienkam sme implementovali aj UI, ktoré nebolo povinnou časťou. Finálnym výstupom je plne funkčná webová aplikácia, kde si môže každý prihlásený používateľ vytvárať vlastné projekty a nastavovať parametre podľa vlastného uváženia. Výstupom pre užívateľov, ktorí nie sú oprávnený vstupovať do aplikácie a ich e-mailové adresy špecifikoval admin pri vytvorení projektu, je report v .csv formáte s odkazom na články danej témy, usporiadané podľa získaného skóre. Podobne ako v zimnom semestri aj teraz sme rozdelenie úloh zaznamenávali v zápisniciach, ktoré sú priložené v nasledujúcej kapitole. Aj počas letného semestra sa nám podarilo splniť stanovené ciele.

Zápisnice

Jedným z úloh tímového projektu bolo aj vypracovanie zápisníc z pravidelných stretnutí tímu. Náš tím mal dva druhy stretnutí:

- s týždenným intervalom - tohto stretnutia sme sa zúčastnili len členovia tímu, aby sme si medzi sebou rozdelili úlohy a povedali si, ako prebiehajú stanovené úlohy.
- dvojtýždňovým intervalom - na tomto stretnutí sa zúčastňoval celý tím aj s vedúcim projektu. Vedúcemu projektu sme vždy zreferovali, v akom stave sa projekt nachádza, aké sú ďalšie plány a podobne. Taktiež sme s ním konzultovali nasledovné ciele a úlohy.

V posledných týždňoch letného semestra sa konali stretnutia aj dvakrát do týždňa, pretože bolo treba častejšie konzultovať ukončenie projektu.

Všetky stretnutia prebiehali online formou prostredníctvom aplikácie MS Teamse. Na nasledujúcich stranách sú zobrazené zápisnice zo zimného aj letného semestra.

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 21.9.2021 17:00

Miesto stretnutia: MS Teams

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 1.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

TÉMY STRETNUTIA

Téma č. 1: Vypracovanie ponuky pre zadávateľa projektu

Popis: S tímom sme si prešli jednotlivé body, ktoré má ponuka obsahovať. Následne sme si zadali nejaké úlohy, ktoré je potrebné vypracovať a zahrnúť do ponuky pre zadávateľa projektu.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Predstavenie členov tímu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	24.9.2021
2.	Spísanie motivácie	Andrej Mikuš	24.9.2021
3.	Vytvorenie formuláru s rozvrhmi	Zuzana Stančíková	24.9.2021

Téma č. 2: Tvorba webovej stránky

Popis: Medzi členov tímu sme si rozdelili úlohy, ktoré je potrebné vykonať na vytvorenie webovej stránky nášho tímu.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Návrh web stránky	Martina Krnáčová, Adam Klimko, Ľubomíra Števuliaková	24.9.2021
3.	Tvorba web stránky	Martina Krnáčová, Adam Klimko, Ľubomíra Števuliaková	24.9.2021

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 21.9.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 27.9.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 2.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Predstavenie členov tímu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	24.9.2021	SPLNENÉ
2.	Spísanie motivácie	Andrej Mikuš	24.9.2021	SPLNENÉ
3.	Vytvorenie formuláru s rozvrhmi	Zuzana Stančíková	24.9.2021	SPLNENÉ
4.	Návrh web stránky	Martina Krnáčová, Adam Klimko, Ľubomíra Števuliaková	24.9.2021	SPLNENÉ
5.	Tvorba web stránky	Martina Krnáčová, Adam Klimko, Ľubomíra Števuliaková	24.9.2021	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Naplánovanie stretnutia s vedúcim témy.

Popis: S tímom sme sa dohodli, že potrebujeme naplánovať stretnutie s vedúcim témy, kde by sme prebrali bližšiu špecifikáciu zadania.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Naplánovať meeting	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	1.10.2021

Téma č. 2: Bližšie oboznámenie sa s témou zadania

Popis: S tímom sme sa dohodli, že každý člen tímu si bližšie naštuduje tému projektu a na ďalšom stretnutí skonzultujeme, kto a čo zistil, aby sme to následne vedeli pridať do dokumentácie.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Naštudovať tému	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	1.10.2021

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 1.10.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 6.10.2021

Miesto stretnutia: MS Teams

Autor zápisu: Andrej Mikuš

Poradové číslo zápisnice: 3.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Petráš Michal

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Naplánovať meeting	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	1.10.2021	SPLNENÉ
2.	Naštudovať tému	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	1.10.2021	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Bližšie oboznámenie s projektom.

Popis: S tímom sme sa stretli s vedúcim projektu, ktorý nás bližšie oboznámil s detailným popisom témy a jednotlivými požiadavkami. Zodpovedali sa všetky otázky a nejasnosti. Taktiež sa stanovila predbežná metodika prác na projekte.

Téma č. 2: Rozdelenie tímu

Popis: Vedúci práce nám zadal úlohu pre rozdelenie nášho tímu na dve alebo tri skupiny a to:

- Skupina so zameraním na analýzu dátových zdrojov.
- Skupina s technickým zameraním na NLP.
- Skupina pozostávajúca z predošlých dvoch(dobrovoľná).

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Rozdelenie tímu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	13.10.2021

Téma č. 3: Získať informácie pre upresnenie plánu.

Popis: Kontaktovať pedagogického vedúceho ohľadom očakávaných výstupov pre jednotlivé míľniky a oboznámiť technického vedúceho. Ďalej upresniť detailný plán v časovom horizonte do záverečnej obhajoby.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Upresnenie detailného plánu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	13.10.2021
2.	Zistiť očakávané výstupy	Marek Mlynček	13.10.2021

Téma č. 4: Nastaviť komunikačné rozhrania s vedúcim práce.

Popis: Vytvoriť spoločnú skupinu na vyhovujúcom médiu a stanoviť pravidelný čas stretnutí s technickým vedúcim práce.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Výber ideálneho času stretnutí.	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	13.10.2021
2.	Nastavenie komunikačného média.	Marek Mlynček	13.10.2021

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 6.10.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 13.10.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 3.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števeliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Rozdelenie tímu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števeliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	13.10.2021	SPLNENÉ
2.	Upresnenie detailného plánu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števeliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	13.10.2021	SPLNENÉ
3.	Zistiť očakávané výstupy	Marek Mlynček	13.10.2021	SPLNENÉ
4.	Výber ideálneho času stretnutí.	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števeliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	13.10.2021	SPLNENÉ
5.	Nastavenie komunikačného média.	Marek Mlynček	13.10.2021	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Vyhodnotenie úloh z predchádzajúceho stretnutia.

Popis: S tímom sme konzultovali, či sme splnili všetky úlohy za posledný týždeň.

Téma č. 2: Štúdium open source platforiem Keras, TensorFlow

Popis: Každý člen tímu si má naštudovať platformy Keras a TensorFlow.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Štúdium platforiem	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	20.10.2021

Téma č. 3: Analýza web portálov

Popis: Každý člen má za úlohu si prezrieť web portály a zistiť, ktoré webové portály obsahujú údaje, ktoré by nám pomohli k riešeniu nášho zadania. S vedúcim projektu následne skonzultujeme zistené informácie.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Analýza web portálov	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	20.10.2021

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 15.10.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 20.10.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 5.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Štúdium platforiem	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	20.10.2021	SPLNENÉ
2.	Analýza web portálov	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	20.10.2021	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Analýza web portálov

Popis: Tento týždeň sa nám nepodarilo naplánovať si stretnutie s vedúcim témy, pretože bol zaneprázdnený, a teda sme s ním nemohli skonzultovať zistené informácie. Aspoň s tímom sme si prešli zistené informácie o webových portáloch.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Naplánovať meeting s vedúcim projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	25.10.2021

Téma č. 2: Tvorba tréningového dátového setu kľúčových slov

Popis: Analýza vhodnej štruktúry a formy dátového setu pre NLP model. Analýza vhodných atribútov pre model, ktoré sú k dispozícii na väčšine zdrojových portálov.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Tvorba dátového setu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	27.10.2021

Téma č. 3: Dokumentácia

Popis: Tvorba dokumentácie

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	Do konca zimného semestra

Téma č. 4: Štúdium NLP

Popis: Analýza známych existujúcich NLP modelov a prístupov k trénovaniu. Príprava softvérových nástrojov, knižníc a prostredia pre tvorbu vlastného modelu.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Štúdium NLP	Adam Klimko, Andrej Mikuš, Marek Mlynček	27.10.2021

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 22.10.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 27.10.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 6.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Naplánovať stretnutie s vedúcim projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	25.10.2021	SPLNENÉ
2.	Tvorba dátového setu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	27.10.2021	NESPLNENÉ
3.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	ROZPRACOVANÉ
4.	Štúdium NLP	Adam Klimko, Andrej Mikuš, Marek Mlynček	27.10.2021	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Analýza web portálov

Popis: S vedúcim témy sme si prešli zistené informácie o web portáloch. Väčšina webových portálov nebola vhodná, pretože sa v nich nenachádzali potrebné kľúčové slová alebo články boli neaktuálne. Z toho dôvodu sme sa s vedúcim projektu dohodli, že tím Analytikov a zmiešaný tím skúsi nájsť nejaké iné webové portály. Vedúci projektu nám v e-maili poslal odvetvia, na ktoré by sme sa mali zamerať. Nájsené portály opäť prekonzultujeme na stretnutí s vedúcim.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Analýza web portálov	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková	10.11.2021

Téma č. 2: Tvorba dátového setu

Popis: Keďže sa nám úlohu nepodarilo vypracovať v priebehu týždňa, predlžujeme termín o týždeň.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Tvorba dátového setu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	3.11.2021

Téma č. 3: Naplánovanie pravidelných stretnutí s vedúcim projektu

Popis: S vedúcim projektu sme sa dohodli na pravidelných stretnutiach, ktoré sa budú konať každý druhý týždeň odo dňa stretnutia. Okrem toho sa my ako tím budeme naďalej stretávať každý týždeň, kde si povieme krátke zhrnutie, čo sme za posledný týždeň stihli alebo nestihli urobiť.

Téma č. 4: Voľba projektového manažéra

Popis: Jeden človek z tímu sa stane projektovým manažérom, ktorý bude mať za úlohu vytvárať zápisnice z každého stretnutia, zadávať každému členovi tímu úlohy, zaznamenávať úlohy cez nejaký Task Planner...

Téma č. 5: Tvorba modelu na vyhľadávanie stránok

Popis: Pomocou python knižnice google/urllib tím NLP, ak bude treba aj zmiešaný tím, pripraví demo verziu modelu na vyhľadávanie vhodných webových stránok pomocou zadaných kľúčových slov.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Tvorba modelu na vyhľadávanie stránok	Adam Klimko, Andrej Mikuš, Marek Mlynček, Martina Krnáčová, Ľubomíra Števuliaková	10.11.2021

Téma č. 6: Tvorba písomného výstupu (dokumentácia)

Popis: Zuzana Stančíková spolu s Klaudiou Kytošovou začali s tvorbou dokumentácie, ktorá má byť vypracovaná do konca semestra. Dokumentácia zatiaľ obsahuje predstavenie tímu, ponuku, krátky opis technológií, s ktorými budeme na projekte pracovať (postupne budeme technológie pridávať, ak budú pribúdať) a tiež zápisnice za posledných päť týždňov.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 28.10.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 3.11.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 7.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza web portálov	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková	10.11.2021	ROZPRACOVANÉ
2.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	ROZPRACOVANÉ
3.	Tvorba dátového setu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	3.11.2021	ROZPRACOVANÉ
4.	Tvorba modelu na vyhľadávanie stránok	Adam Klimko, Andrej Mikuš, Marek Mlynček	10.11.2021	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Pokrok v zadaných úlohách

Popis: S tímom sme konzultovali, v akom stave sú zadané úlohy - či má niekto s niečím problém a treba mu pomôcť a podobne. Žiadne nové úlohy neboli zadané, keďže väčšina úloh neboli dokončené.

Téma č. 2: Tvorba dátového setu

Popis: Opäť sme museli predĺžiť termín na tvorbu dátového setu, nakoľko sme na tvorbu chceli použiť model na vyhľadávanie stránok.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 6.11.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 11.11.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 8.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza web portálov	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková	10.11.2021	HOTOVO
2.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	ROZPRACOVANÉ
3.	Tvorba dátového setu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	11.11.2021	HOTOVO
4.	Tvorba dema na vyhľadávanie stránok	Adam Klimko, Andrej Mikuš, Marek Mlynček	10.11.2021	HOTOVO

TÉMY STRETNUTIA

Téma č. 1: Tvorba dátového setu a tvorba dema na vyhľadávanie stránok

Popis: Prvotná verzie datasetu aj modelu bola vytvorená do 11.11.2021. Demo aj dataset je však potrebné doladiť, pretože nefunguje tak, ako by sme chceli.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Tvorba a úprava dátového setu 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Neuvádza sa
2.	Tvorba a úprava dema na vyhľadávanie stránok	Andrej Mikuš	Neuvádza sa

Téma č. 2: Analýza databáz

Popis: Je potrebné urobiť analýzu dostupných databáz, do ktorých sa budú ukladať exporty z dema. Databázy vhodné na analýzu: Mongo DB, Elastic, MySQL.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Analýza databáz	Adam Klimko	24.11.2021

Téma č. 3: Zhromaždenie neplnovýznamových slovenských slov

Popis: Potrebujeme vytvoriť osobitný dataset, ktorý bude obsahovať slovenské neplnovýznamové slová. Preto treba nájsť a zhromaždiť čo najviac neplnovýznamových slov. Tieto slová sa použijú v skripte, ktorý vyhľadáva kľúčové slová, aby takéto slová pri vyhľadávaní vynechal.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Zhromaždenie neplnovýznamových slov v slovenskom jazyku	Martina Krnáčová	18.11.2021

Téma č. 4: Naplánovať schôdzku s Ing. Juraj Kačur, PhD.

Popis: Potrebujeme sa spojiť a naplánovať si stretnutie s Jurajom Kačúrom. Na stretnutí mu porozprávame o našom projekte a mal by nám dať nejaké rady pri práci na projekte.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Naplánovanie stretnutia	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	24.11.2021

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 11.11.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 18.11.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 9.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza databáz	Adam Klimko	24.11.2021	ROZPRACOVANÉ
2.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	ROZPRACOVANÉ
3.	Zhromaždenie neplnovýznamových slov v slovenskom jazyku	Martina Krnáčová	18.11.2021	HOTOVO
4.	Naplánovanie stretnutia	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	24.11.2021	ROZPRACOVANÉ
5.	Tvorba a úprava dátového setu 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Neuvádza sa	ROZPRACOVANÉ
6.	Tvorba a úprava dema na vyhľadávanie stránok	Andrej Mikuš	Neuvádza sa	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Rozpracované úlohy

Popis: S tímom sme si prešli všetky rozpracované úlohy, povedali sme si aký je pokrok, či niekto s niečím nepotrebuje pomôcť a podobne.

Téma č. 2: Manuálne overovanie stránok

Popis: Je potrebné manuálne kontrolovať stránky, či demo správne analyzuje stránky a nájde požadované kľúčové slová.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Overovanie stránok	Kludia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števeliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do finálnej verzie dema

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 19.11.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 24.11.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 10.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza databáz	Adam Klimko	24.11.2021	HOTOVO
2.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	ROZPRACOVANÉ
3.	Overovanie stránok	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do finálnej verzie dema	ROZPRACOVANÉ
4.	Naplánovanie stretnutia	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	24.11.2021	ROZPRACOVANÉ
5.	Tvorba a úprava dátového setu 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Neuvádza sa	ROZPRACOVANÉ
6.	Tvorba a úprava dema na vyhľadávanie stránok	Andrej Mikuš	Neuvádza sa	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Plány po zimnom semestri

S vedúcim projektu sme sa predbežne snažili naplánovať nasledujúce stretnutia po Novom roku. Všetko však bude závisieť od skúškového obdobia. Na presnom dátume sa pravdepodobne dohodneme až na stretnutí 8.12.2021.

Téma č. 2: Informácia o TP1 prezentáciách.

Popis: Potrebujeme zistiť informácie o prezentáciách, ktoré majú prebiehať po skončení zimného semestra.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Info k prezentáciám	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	8.12.2021

Téma č. 3: Zhromaždenie neplnovýznamových slov slovenského jazyka 2. časť

Popis: Je potrebné zhromaždiť čo najviac neplnovýznamových slov slovenského jazyka, ako aj zámená a číslovky. Pôvodná verzia nestačila.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Zhromaždenie neplnovýznamových slov slovenského jazyka 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková	8.12.2021

Téma č. 4: Pokračovanie práce na rozpracovaných úlohách.

Popis: Na stretnutí sme si prešli, v akej fáze sa projekt nachádza a čo treba doladiť. Potrebujeme upraviť skórovanie stránok, vytvoriť black a white list stránok, pridať spomínané neplnovýznamové slová...

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 25.11.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 4.12.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 11.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza databáz	Adam Klimko	24.11.2021	HOTOVO
2.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	ROZPRACOVANÉ
3.	Overovanie stránok	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do finálnej verzie dema	ROZPRACOVANÉ
4.	Naplánovanie stretnutia	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	24.11.2021	ROZPRACOVANÉ
5.	Tvorba a úprava dátového setu 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Neuvádza sa	ROZPRACOVANÉ
6.	Tvorba a úprava dema na vyhľadávanie stránok	Andrej Mikuš	Neuvádza sa	ROZPRACOVANÉ
7.	Info k prezentáciám	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	8.12.2021	ROZPRACOVANÉ
8.	Zhromaždenie neplnovýznamových slov slovenského jazyka 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková	8.12.2021	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Rozpracované úlohy

Popis: Na poslednom stretnutí sme si len prešli rozpracované úlohy. Z dôvodu časovej tiesne kvôli zápočtom a odovzdávaniu zadaní z viacerých predmetov, sme nedopĺňali nové úlohy.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 5.12.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 10.12.2021

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 12.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
2.	Dokumentácia	Zuzana Stančíková, Klaudia Kytošová	Do konca zimného semestra	HOTOVO
3.	Overovanie stránok	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do finálnej verzie dema	ROZPRACOVANÉ
5.	Tvorba a úprava dátového setu 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Neuvádza sa	ROZPRACOVANÉ
6.	Tvorba a úprava dema na vyhľadávanie stránok	Andrej Mikuš	Neuvádza sa	ROZPRACOVANÉ
7.	Info k prezentáciám	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	8.12.2021	HOTOVO
8.	Zhromaždenie neplnovýznamových slov slovenského jazyka 2. časť	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková	8.12.2021	HOTOVO

TÉMY STRETNUTIA

Téma č. 1: Rozpracované úlohy

Popis: S celým tímom a vedúcim projektu sme si prešli všetky úlohy za posledné dva týždne. Každý z tímu povedal, čo sa mu podarilo resp. nepodarilo. Projekt je celkovo vo veľmi dobrom stave. Je

rozpracované demo na vyhľadávanie a hodnotenie stránok, ktoré sa postupne doladuje. Ďalej začala príprava databázy, z ktorej bude možné vytvárať reporty.

Téma č. 2: Plány na nasledujúce obdobie

Popis: Predbežne sme sa dohodli na nasledujúcom stretnutí, ktoré bude až v januári. Tam by sme si prešli nasledujúce kroky.

Téma č. 3: Kontrola dokumentácie

Popis: Každý člen tímu má za úlohu skontrolovať dokumentáciu, aby sa mohla odovzdať do AIS. V prípade nejakých pripomienok by sa dokumentácia upravila, poprípade by sa doplnili ďalšie nové poznatky.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 10.12.2021

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 14.2.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 1.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Keďže toto je prvá zápisnica v letnom semestri, žiadne predchádzajúce úlohy neboli zadane. V tom prípade nemáme čo vyhodnotiť.

TÉMY STRETNUTIA

Téma č. 1: Naplánovanie stretnutia s vedúcim témy.

Popis: Keďže začal letný semester, bolo potrebné naplánovať pravidelné stretnutia s vedúcim témy.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Naplánovať meeting	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	18.2.2022

Téma č. 2: Info ohľadom stave projektu

Popis: S tímom sme konzultovali, kde sme s projektom skončili, čo už máme hotové a čo nás ešte čaká.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 18.2.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 21.2.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 2.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Naplánovať meeting	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	18.2.2021	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Info k aktuálnemu stavu projektu.

Popis: S vedúcim projektu sme diskutovali, v akom stave je projekt. Rozprávali sme, kde a s čím sme skončili v zimnom semestri.

Téma č. 2: Plány na letný semester

Popis: Vedúci projektu nám zadal niekoľko úloh, na ktorých budeme pracovať počas letného semestra.

Číslo úlohy	Názov úlohy	Popis úlohy	Zodpovedné osoby	Termín
1.	Analýza alternatív k nášmu projektu	Prehľadať na internete alternatívy, či už existujú technológie, ktoré by vedeli robiť to, na čo sa zameriava tento projekt.	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	7.3.2022
2.	Dokumentácia	V dokumentácii lepšie popísať NLP technológiu, doplniť využitie machine learningu	Klaudia Kytošová, Zuzana Stančíková	-
3.	Sekvenčný diagram	Nakresliť a pridať do dokumentácie sekvenčný diagram.	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-

4.	Navrhnuť front-end	Ak zostane čas a priestor, pokúsiť sa navrhnuť front-end, prípadne ho zapracovať do projektu.	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-
----	--------------------	---	--	---

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 25.2.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 4.3.2022

Miesto stretnutia: MS Teams

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 3.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza alternatív k nášmu projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	7.3.2022	ROZPRACOVANÉ
2.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
3.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Plnenie úloh

Popis: S tímom sme si prešli úlohy, ktoré máme zadané a ich aktuálny stav. Ďalej sme potrebovali spracovať otázky ohľadom návrhu front-end aplikácie, ktoré potrebujeme prejsť pri najbližšom calle s vedúcim témy.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1	Spracovať otázky pre vedúceho témy	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	8.3.2022

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 6.3.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 7.3.2022

Miesto stretnutia: MS Teams

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 4.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza alternatív k nášmu projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	7.3.2022	ROZPRACOVANÉ
2.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
3.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
5.	Spracovať otázky pre vedúceho témy	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	8.3.2022	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Poskytnutie servera

Popis: S vedúcim projektu sme riešili poskytnutie servera pre náš projekt, prostredníctvom ktorého by si každý člen tímu mohol pozrieť, v akom stave je projekt a ako funguje. Potrebujeme teda zistiť, či vie škola zabezpečiť nejaký server, na ktorý by sa demo nasadilo.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Napísať vyučujúcim email ohľadom servera	Zuzana Stančíková	21.3.2022

Téma č. 2: Front-end z pohľadu vedúceho.

Popis: Zodpovedanie našich otázok ohľadom front-endu. Appka by mala obsahovať: vytvorenie účtu, prihlasovanie užívateľov, možnosť vytvoriť si projekt, tím užívateľov... Z témy nevyplynuli žiadne ďalšie úlohy.

Téma č.3: Kategorizácia článkov a tvorba slovníka

Popis: Premyslieť ako budeme kategorizovať články. Rozšírenie slovníkov, váhovanie slov...

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Slovníky	Klaudia Kytošová, Zuzana Stančíková, Marek Mlynček, Andrej Mikuš	-

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 8.3.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 14.3.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 5.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza alternatív k nášmu projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	7.3.2022	ROZPRACOVANÉ
2.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
3.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
5.	Slovníky	Klaudia Kytošová, Zuzana Stančíková, Marek Mlynček, Andrej Mikuš	21.3.2022	ROZPRACOVANÉ
6.	Napísať vyučujúcim email ohľadom servera	Zuzana Stančíková	21.3.2022	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Poskytnutie virtualizačnej služby

Popis: Je potrebné žiadosťou kontaktovať vedúceho VS, Ing. Miroslav Peško, ktorý má na starosti poskytovanie virtuálnych serverov. S vytvorením virtuálneho servera by nám mal pomôcť Ing. Štefan Balogh PhD.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Podat' žiadosť na poskytnutie virtualizačnej služby	Klaudia Kytošová	21.3.2022

Téma č. 2: Zhodnotenie rozpracovaných úloh

Popis: S tímom sme si prešli rozpracované úlohy, v akom sú stave a či nie je s niečím problém.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam. Počas stretnutia nevznikli žiadne ďalšie úlohy.

Dátum vytvorenia zápisnice: 20.3.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 14.3.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 5.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza alternatív k nášmu projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	7.3.2022	ROZPRACOVANÉ
2.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
3.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
5.	Slovníky	Klaudia Kytošová, Zuzana Stančíková, Marek Mlynček, Andrej Mikuš	21.3.2022	ROZPRACOVANÉ
6.	Napísať vyučujúcim email ohľadom servera	Zuzana Stančíková	21.3.2022	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Poskytnutie virtualizačnej služby

Popis: Je potrebné žiadosťou kontaktovať vedúceho VS, Ing. Miroslav Peško, ktorý má na starosti poskytovanie virtuálnych serverov. S vytvorením virtuálneho servera by nám mal pomôcť Ing. Štefan Balogh PhD.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Podat' žiadosť na poskytnutie virtualizačnej služby	Klaudia Kytošová	21.3.2022

Téma č. 2: Zhodnotenie rozpracovaných úloh

Popis: S tímom sme si prešli rozpracované úlohy, v akom sú stave a či nie je s niečím problém.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam. Počas stretnutia nevznikli žiadne ďalšie úlohy.

Dátum vytvorenia zápisnice: 20.3.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 21.3.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 6.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Analýza alternatív k nášmu projektu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	7.3.2022	ROZPRACOVANÉ
2.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
3.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
5.	Slovníky	Klaudia Kytošová, Zuzana Stančíková, Marek Mlynček, Andrej Mikuš	21.3.2022	SPLNENÉ
6.	Podat' žiadosť na poskytnutie virtualizačnej služby	Klaudia Kytošová	21.3.2022	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: S vedúcim tímu sme si prešli aktuálny stav projektu. Čo sa nám podarilo za dva týždne urobiť. Podarilo sa:

- Váhovanie slov v slovníku, filtrácia nepotrebných JS slov
- Čiastočný návrh front-endu

- Podanie žiadosti na poskytnutie servera
- Zapracovanie e-mailových notifikácií o generovaní reportu

Z témy vyplynula nová úloha, ktorou je manuálne kontrolovanie článkov, ktoré vygeneruje report.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Manuálna kontrola článkov	Klaudia Kytošová, Zuzana Stančíková	4.4.2022

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 21.3.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 28.3.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 7.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
2.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
3.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Manuálna kontrola článkov	Klaudia Kytošová, Zuzana Stančíková	4.4.2022	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Na stretnutí sme si prešli všetky rozpracované úlohy. Na rozpracovaných úlohách sa postupuje podľa plánu a nevznikajú žiadne problémy. Nevznikli žiadne nové úlohy. Jedine potrebujeme doriešiť zapožičanie serveru zo školy.

Téma č. 2: Plánovanie teambuildingu

Popis: Na utuženie vzájomných vzťahov sme sa rozhodli naplánovať teambuilding. 😊

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 3.4.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 4.4.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 8.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
2.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
3.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Manuálna kontrola článkov	Klaudia Kytošová, Zuzana Stančíková	4.4.2022	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Na stretnutí sme si prešli všetky rozpracované úlohy. Na rozpracovaných úlohách sa postupuje podľa plánu a nevznikajú žiadne problémy. Taktiež nevznikajú žiadne nové veľké úlohy, nakoľko pomaly finišujeme s projektom.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Zistiť info ohľadom odovzdávania TP	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	11.4.2022
2.	Zapožičanie serveru	Klaudia Kytošová	18.4.2022

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 10.4.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 11.4.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 9.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
2.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
3.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
4.	Zapožičanie serveru	Klaudia Kytošová	18.4.2022	SPLNENÉ
5.	Zistiť info ohľadom odovzdávania TP	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	11.4.2022	SPLNENÉ

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Na stretnutí sme si prešli všetky rozpracované úlohy. Nevznikli žiadne nové úlohy. Dohodli sme sa, že najbližšie stretnutie bude až 25.4.2022 kvôli veľkonočným sviatkom.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 18.4.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 21.4.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 10.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	-	ROZPRACOVANÉ
2.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ
3.	Navrhnuť front-end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	HOTOVO

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Kvôli veľkonočným sviatkom a zvýšenému počtu zápočtov / zadaní sme stretnutie naplánovali až na 11. týždeň aj s vedúcim témy. Na projekte sa však usilovne pracuje podľa plánu.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 21.4.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 25.4.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 11.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	Do konca semestra	ROZPRACOVANÉ
2.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	ROZPRACOVANÉ

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Pôvodne sme tento týždeň mali mať stretnutie aj s vedúcim témy. Stretnutie sme však museli nečakane presunúť na ďalší týždeň, pretože sa vedúci nemohol stretnutia zúčastniť. V rámci tímu sme si však na stretnutí povedali, čo musíme urobiť do konca semestra. Potrebujeme dokončiť dokumentáciu, spojiť back end s front end, nasadiť projekt na školský server, natočiť video prezentáciu. S vedúcim témy si musíme dohodnúť stretnutie, aby sme mu vedeli povedať, ako sme na tom s projektom. Taktiež sa s ním musíme dohodnúť, čo od nás do konca semestra ešte očakáva.

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín
1.	Back end + front end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra
2.	Nasadiť projekt na server	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra

3.	Natočiť video prezentáciu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra
----	---------------------------	--	-------------------

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 29.4.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 5.5.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 12.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	Do konca semestra	ROZPRACOVANÉ
2.	Sekvenčný diagram	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	-	HOTOVO
3.	Back end + front end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra	ROZPRACOVANÉ
4.	Nasadiť projekt na server	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra	ROZPRACOVANÉ
5.	Natočiť video prezentáciu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra	-

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Tento týždeň sa nám podarilo zorganizovať stretnutie aj s vedúcim projektu. Oboznámili sme ho s aktuálnym stavom projektu. Momentálne prebieha nasadenie projektu na server a posledné doladovanie backend a frontend. Dokumentácia je už skoro hotová, treba doplniť len zopár UML diagramov, popísať implementačnú časť a opis výsledkov. Z dnešného stretnutia nevyplývajú skoro

žiadne úlohy. Má sa dokončiť všetko, čo je rozpracované. Vedúcemu sa do konca týždňa pošle zatiaľ nedokončená dokumentácia na čiastočnú kontrolu, aby sme ju vedeli v 13. týždni doladiť a dokončiť.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 5.5.2022

ZÁPIS ZO STRETNUTIA

Tímový projekt 2021/2022

Názov projektu: Získavanie a využitie informácií z verejne dostupných zdrojov s podporou open source technológií Machine Learning a AI

Dátum a čas stretnutia: 9.5.2022

Miesto stretnutia: MS Teamse

Autor zápisu: Klaudia Kytošová

Poradové číslo zápisnice: 13.

Prítomní: Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček, Michal Petráš

VYHODNOTENIE ÚLOH Z PREDCHADZAJÚCEHO STRETNUTIA

Číslo úlohy	Názov úlohy	Zodpovedné osoby	Termín	Stav
1.	Dokumentácia	Klaudia Kytošová, Zuzana Stančíková	Do konca semestra	ROZPRACOVANÉ
2.	Back end + front end	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra	HOTOVO
3.	Nasadiť projekt na server	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	Do konca semestra	HOTOVO
4.	Natočiť video prezentáciu	Klaudia Kytošová, Martina Krnáčová, Zuzana Stančíková, Ľubomíra Števuliaková, Adam Klimko, Andrej Mikuš, Marek Mlynček	15.5.2022	-

TÉMY STRETNUTIA

Téma č. 1: Zhodnotenie rozpracovaných úloh

Popis: Od posledného stretnutia sa projekt úspešne podarilo nasadiť na server. Taktiež nám vedúci projektu skontroloval doterajšiu dokumentáciu. Je potrebné v nej urobiť zopár zmien (poprehadzovať kapitoly...). Do odovzdania projektu budeme mať ešte jedno stretnutie, na ktorom zhodnotíme dokončenú dokumentáciu a vedúcemu ukážeme video prezentáciu. Zo stretnutia nevyplynuli žiadne nové úlohy.

Záver: Na stretnutí sa zúčastnili všetci členovia tímu. Všetci členovia tímu sa aktívne zapájali do diskusie k jednotlivým témam.

Dátum vytvorenia zápisnice: 9.5.2022

Zoznam použitej literatúry

1. Dostupné tiež z: <https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP/>.
2. Dostupné tiež z: <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML>.
3. Dostupné tiež z: <https://techvidvan.com/tutorials/advantages-and-disadvantages-of-machine-learning/>.
4. RAJPUT, Sambhav. *What are the reasons to choose Laravel MVC for web development*. 2020.
5. BHATT, Shardul. *Pros and Cons of Django Framework for App Development*. <https://dzone.com/articles/pros-and-cons-of-django-framework-for-app-developm>.
6. MIHALCHENKO, Anatoly. *Pros and Cons of Using Spring Boot*. <https://scand.com/company/blog/pros-and-cons-of-using-spring-boot/>.
7. ŠTEVULIAKOVÁ, Lubomíra. *Využitie moderných webových technológií pri návrhu interetového obchodu alternatívnych učebných materiálov pre ZŠ*. 2022.
8. Dostupné tiež z: <https://www.xplenty.com/blog/which-database/>.
9. Dostupné tiež z: <https://www.xplenty.com/blog/the-sql-vs-nosql-difference/>.
10. Dostupné tiež z: <https://www.geeksforgeeks.org/mongodb-vs-mysql/>.
11. Dostupné tiež z: <https://www.elastic.co/what-is/kibana>.
12. Dostupné tiež z: <https://www.javatpoint.com/advantages-and-disadvantages-of-elasticsearch>.
13. *What is REST?* <https://www.codecademy.com/article/what-is-rest>.
14. BAELDUNG. *Learn JPA & Hibernate*. 2021. <https://www.baeldung.com/learn-jpa-hibernate>.
15. *Diffbot*. <https://www.diffbot.com/>.
16. *Diffbot wiki*. <https://en.wikipedia.org/wiki/Diffbot>.
17. *The Best Web Scraping Tools for 2021*. <https://www.scrapingbee.com/blog/web-scraping-tools/>.

18. *ScrapeBox*. <http://www.scrapebox.com/>.

Prílohy

A	Api dokumentácia	II
---	----------------------------	----

A Api dokumentácia

API dokumentácia k webovej aplikácii. Ak nie je uvedené inak content-type je application/json a parametre sú povinné.

URI: POST /api/registration

- **Opis:** Vytvorí nového používateľa.
- **Request body:**
 - username:** unikátne meno [String]
 - password:** heslo [String]
 - email:** unikátny email [String]
- **Responses**
 - 200 ok
 - 422 Unprocessable entity | Meno a email musia byť unikátne
 - 404 Not Found
 - 400 Bad Request

URI: POST /api/login

- **Opis:** Autentifikácia a priradenie [jwt].
- **Request body:**
 - username:** prihlasovacie meno [String]
 - password:** heslo [String]
- **Responses**
 - 200 ok
 - 400 Bad request
 - 403 Forbidden | Používateľ s menom [username] neexistuje

URI: PUT /api/user

- **Opis:** Upraví zadaný údaj o používateľovi.
- **Request header:**
 - Authorization:** JWT [String]

- **Request parametre:**
 - username:** unikátne meno (nepovinné) [String]
 - password:** heslo (nepovinné) [String]
 - email:** unikátny email(nepovinné) [String]
- **Responses** 200 ok | UserDto
 - 404 Not Found
 - 403 Forbidden | Používateľ s menom [username] neexistuje
 - 400 Bad request| Zadané údaje niesú validné

URI: POST /api/project

- **Opis:** Pridá nový projekt do databázy, vytvorí príslušné súbory s pustí CronJob pre generovanie reportov
- **Request header:**
 - Authorization:** JWT [String]
- **Request body:**
 - title:** názov projektu [String]
 - webpages:** zoznam url stránok [List<String>]
 - forbiddenWebpages:** zoznam url stránok [List<String>]
 - usersEmail:** zoznam emailov [List<String>]
 - dateAfter:** hodnota pre opakovanie generovania [DateAfterEnum(DAY,WEEK,MONTH,YEAR)]
 - dict:** ohodnotený slovník [List<DictionaryItemDto>]
 - keywords:** zoznam kľúčových slov [List<String>]
- **Responses** 200 ok UUID
 - 403 Forbidden
 - 400 Bad request | Zadané údaje niesú validné
 - 500 Internal Server Error

URI: PUT /api/project/id

- **Opis:** Upraví existujúci projekt s daným id
- **Request header:**
 - Authorization:** JWT [String]

- **Request body:**
 - title:** názov projektu [String]
 - active:** aktivnosť [boolean]
 - webpages:** zoznam url stránok [List<String>]
 - forbiddenWebpages:** zoznam url stránok [List<String>]
 - usersEmail:** zoznam emailov [List<String>]
 - dateAfter:** hodnota pre opakovanie generovania [DateAfterEnum(DAY,WEEK,MONTH,YEAR)]
 - dict:** ohodnotený slovník [List<DictionaryItemDto>]
 - keywords:** zoznam kľúčových slov [List<String>]
- **Responses**
 - 200 ok
 - 403 Forbidden
 - 400 Bad request | Zadané údaje nie sú validné
 - 404 Not found | Projekt s id neexistuje
 - 500 Internal Server Error

URI: GET /api/projects

- **Opis:** Vrátí všetky projekty, ktoré založil prihlásený používateľ
- **Request header:**
 - Authorization:** JWT [String]
- **Responses**
 - 200 ok List<ProjectListItemDto>
 - 403 Forbidden

URI: GET /api/projects/id

- **Opis:** Vrátí detail projektu s daným id, ktorý založil prihlásený používateľ
- **Request header:**
 - Authorization:** JWT [String]
- **Responses**
 - 200 ok ProjectDto
 - 403 Forbidden

URI: DELETE /api/projects/id

- **Opis:** Zmaže projekt s daným id, ktorý založil prihlásený používateľ. Zmazanie sa vzťahuje na odstránenie záznamu v databáze, zmazanie príslušných súborov a zrušenie CronJobu
- **Request header:**
Authorization: JWT [String]
- **Responses** 200 ok ProjectDto
404 Not found | Projekt s id neexistuje
403 Forbidden