

Tímový projekt
Projektová dokumentácia

Členovia tímu: Jakub Hlavačka, Dominik Horváth, Jakub Müller, Táňa Poláková, Dávid Silady, Adam Šípka

Vedúci tímu: Ing. Richard Marko, PhD.

Študijný program: Inteligentné softvérové systémy

Ročník: 1.

Kontakt na tím: tim8.fiit.stuba@gmail.com

Inžinierske dielo

Dokumentácia

Členovia tímu: Jakub Hlavačka, Dominik Horváth, Jakub Müller, Táňa Poláková, Dávid Silady, Adam Šípka

Vedúci tímu: Ing. Richard Marko, PhD.

Študijný program: Inteligentné softvérové systémy

Ročník: 1.

Kontakt na tím: tim8.fiit.stuba@gmail.com

Big Picture	4
Úvod	4
Globálne ciele projektu na zimný semester	5
Vzorka dát	5
Čistenie dát	5
Uloženie dát	5
Webová stránka tímu	5
Indexovanie dát	5
Funkčné webové rozhranie pre AMS	6
Funkčná API	6
Prototyp so statickým sťahovaním a indexovaním	6
Globálne ciele projektu na letný semester	7
Stránkovanie výsledkov	7
Filtrovanie vyhľadávania	7
Účet používateľa	7
Väčšie množstvo dát	7
Automatický zber dát	8
Zobrazenie archivovaných článkov	8
PDF report	8
Hotový produkt	8
Celkový pohľad na systém	9
Vysoká architektúra	9
Diagram prípadov použitia	10
Sekvenčný diagram vyhľadávania v systéme	11
Sekvenčný diagram fungovania webovej aplikácie	12
Sekvenčný diagram fungovania API serveru	13
Sekvenčný diagram fungovania scrapovania	14
Dátový diagram PostgreSQL	15
Moduly systému	16
Modul Elasticsearch	16
Analýza	16
Návrh	16
Implementácia	17
Testovanie	17
Modul Aplikačný server	18
Analýza	18
Návrh	18
Implementácia	18
Webové rozhranie	18
Statické súbory	19
Prihlasovanie	19
Databáza	19
Vývojové prostredie	19
Testovanie	19

Modul Frontend	21
Analýza	21
Návrh	21
Implementácia	21
Štýl dizajnu	21
Domovská obrazovka	22
Základné vyhľadávanie	23
Rozšírené vyhľadávanie	24
Stránka o projekte	25
Archív	25
PDF report	26
Správa používateľa	27
Testovanie	27
Modul API server	28
Analýza	28
Návrh	28
Implementácia	28
Testovanie	29
Modul Scraper	30
Analýza	30
Návrh	30
Implementácia	30
Testovanie	31
Prílohy	31
Protokol z testovania	32
Hodnotenie funkcionality advanced search	32
Hodnotenie funkcionality archive	32
Hodnotenie funkcionality pdf report	32
Hodnotenie celkovej práce s prototypom	32
Ďalšie postrehy	33
Technická dokumentácia	34

Big Picture

Úvod

Táto časť dokumentu slúži na predstavenie Inžinierskeho diela, ktoré je aktuálne výstupom našej spoločnej práce na tímovom projekte. V prvej kapitole je opísaný vývoj globálnych cieľov počas zimného semestra. Ďalšia kapitola je venovaná globálnym cieľom v rámci letného semestra. Obsah poslednej kapitoly patrí celkovému pohľadu na systém, ktorý sme sa snažili priblížiť najmä prostredníctvom diagramov.

Globálne ciele projektu na zimný semester

S tým ako plynuli týždne a šprinty v zimnom semestri sa vyvíjali aj globálne ciele tohto projektu. Jednotlivé globálne ciele sú opísané v kapitolách tejto podkapitoly, pričom je dodržaný chronologický postup akým sme ich napĺňali.

Vzorka dát

Naším prvým cieľom bolo získať informácie o údajoch a údajových zdrojoch, ktoré bude náš systém využívať. Ako výborný prameň dát sa ukázal Google RSS, z ktorého sme boli schopní získať našu prvotnú vzorku dát. Táto vzorka bola zložená z článkov o zločinoch v rámci jedného mesiaca. Jej celková veľkosť bola ~30GB a články v nej pokrývali cez 100 zločinov.

Čistenie dát

Vzhľadom na veľkosť samotnej vzorky dát, za časové obdobie jedného mesiaca, sme museli nájsť spôsob, akým tieto dáta zredukovať. Rozhodli sme sa preto vyberať iba potrebný text z nadpisov a paragrafov článku. Úlohou systému má byť indexovanie obsahu samotných článkov, a preto sa tento spôsob osvedčil ako postačujúci.

Ďalej sme sa aplikovali kompresiu Huffmanovho kódovania a výsledkom bolo zredukovanie ~30GB na ~1.5GB bez aplikovania Huffmanovho kódovania a ~350MB po jeho aplikovaní.

Uloženie dát

V analýze pre tento cieľ sme zistili, že jednou z vhodných databáz pre veľké dáta je MongoDB.

Túto databázu sme si vybrali z nasledovných dôvodov:

- Jednoduchá integrácia so Scrapy systémom, ktorý používame na sťahovanie dát.
- Zabudovaná zlíb kompresia.
- Jednoduchá integrácia s Elasticsearch, vďaka JSON formátu samotných dokumentov.

Webová stránka tímu

Po udelení prístupového SSH kľúča, bolo našou prioritou vytvorenie webovej stránky tímu. Táto webová stránka má umožňovať jednoduchý prístup k zápisnicam zo stretnutí, exportom úloh a ostatným dokumentom.

Indexovanie dát

Hlavnou funkcionalitou systému Adverse Media Screening je možnosť vyhľadávania fyzických a právnických osôb spojených so zločini, ktoré údajne vykonali. Aby sme tieto osoby a zločiny dokázali jednoducho vyhľadať, bolo potrebné začať indexovať obsah stiahnutých článkov. V rámci analýzy sme sa pozreli na rôzne indexovacie systémy a nakoniec sme vybrali Elasticsearch.

Funkčné webové rozhranie pre AMS

Predpokladáme, že používatelia, ktorí budú náš systém používať, nemajú znalosti ako jeho vývojári alebo iní kolegovia z IT oblasti. Preto bolo úlohou ďalšieho cieľa vytvoriť užívateľsky priateľské rozhranie, ktoré bude používateľom ľahko prístupné prostredníctvom webovej stránky.

Funkčná API

Námet na vytvorenie aplikačno-programovacieho rozhrania (API) prišiel od nášho vedúceho. Cieľom našej API je sprístupniť možnosť vyhľadávania kriminálnych priestupkov fyzických a právnických osôb aj používateľom, ktorí nebudú používať systém Adverse Media Screening, ale budú chcieť využívať túto funkčnosť napríklad vo svojom vlastnom systéme.

Prototyp so statickým sťahovaním a indexovaním

Výsledkom globálnych cieľov za zimný semester bol prvý funkčný prototyp, prostredníctvom ktorého používateľ dokázal vyhľadávať nad menším počtom článkov. Databáza v tom čase obsahovala ~51000 článkov z obdobia 15.11.2021 - 29.11.2021. Vyhľadávanie je jednoduché, príjemné a intuitívne.

Globálne ciele projektu na letný semester

V tejto kapitole sa bližšie pozrieme na globálne ciele nášho tímu v letnom semestri. Niektoré z nich sme vedeli identifikovať ešte pred začatím semestra, ostatné vznikli počas jeho priebehu. Väčšinou sa jedná o rozšírenú funkcionality už existujúcich častí riešenia.

Stránkovanie výsledkov

Prototyp z predchádzajúceho semestra poskytoval len základné vyhľadávanie pričom vracal fixný počet výsledkov, pričom relevantných článkov mohlo byť aj viac. Preto bolo nutné vyriešiť stránkovanie výsledkov na strane API servera a taktiež aj webového rozhrania, aby používatelia mohli s jednoduchosťou prechádzať cez všetky vrátené články.

Filtrovanie vyhľadávania

V predchádzajúcej verzii AMS bolo možné vyhľadávať len podľa mena fyzickej, či právnickej osoby bez akejkoľvek možnosti filtrovania nájdených výsledkov. Rozhodli sme sa poskytnúť filtrovanie podľa možných zločinov, krajiny pôvodu článku a roku, v ktorom bol napísaný. Jednotlivé zločiny boli neskôr rozdelené do rôznych kategórií, ako napríklad vandalizmus, finančné zločiny, kybernetické zločiny a pod. Tým sme zvýšili prehľadnosť a jednoduchosť používateľského rozhrania. Celkovo sa jednalo o zložitejšiu úlohu, na ktorej sa postupne pracovalo takmer počas celého semestra, keďže vyžadovala doplnenie funkcionality vo viacerých častiach nášho systému.

Účet používateľa

Pôvodná webová aplikácia sa dala používať bez nutnosti registrácie, či prihlasovania. Samotné vyhľadávanie chceme poskytovať všetkým aj naďalej, avšak došli sme k záveru, že niektoré budúce funkcie by mali byť dostupné len pre registrovaných používateľov. Preto boli vytvorené dva formuláre, konkrétne pre registráciu a prihlásenie, ktoré momentálne obsahujú používateľské meno a heslo. Samozrejme v tomto kroku bolo nutné vyriešiť autentifikáciu, zabezpečenie a taktiež aj používateľskú databázu. Usúdili sme, že v tomto kroku je ideálnejšie použiť SQL databázu, preto sme sa rozhodli pre PostgreSQL.

Väčšie množstvo dát

Ako už bolo vyššie spomenuté, pôvodný prototyp pracoval len s malým množstvom článkov pochádzajúcich len z Veľkej Británie z krátkého časového obdobia. Bolo teda nutné získať články zo všetkých krajín Európskej únie za niekoľko posledných rokov. Avšak nie všetky krajiny EÚ boli nakoniec použité, keďže ich média neboli v Google RSS feed-e dostupné. Celkovo sa nám však podarilo získať články pochádzajúce z 21 krajín, od roku 2016 až po súčasnosť. Zoznam jednotlivých krajín je uvedený v nasledovnej tabuľke.

Rakúsko	Belgicko	Bulharsko	Česká republika	Francúzsko	Nemecko	Veľká Británia
Grécko	Maďarsko	Taliansko	Lotyšsko	Litva	Holandsko	Poľsko
Portugalsko	Rumunsko	Slovensko	Slovinsko	Švajčiarsko	Ukrajina	USA

Automatický zber dát

Chceme, aby články boli čo najaktuálnejšie, avšak manuálne spúšťať scraper každý deň je vysoko neefektívne. Preto bolo nutné tento proces automatizovať. Bol vytvorený tzv. plánovač získavania článkov, ktorý každé ráno spustí scraper, pričom získava články z predchádzajúceho dňa zo všetkých nami definovaných regiónov. V aktuálnom stave máme k dispozícii už okolo 2,4 milióna článkov, ktorých veľkosť na disku je približne 80 GB (60 GB pre Elasticsearch, 20 GB pre MongoDB). Avšak tieto čísla neustále stúpajú, keďže celkový nárast článkov je približne 15 000 za každý deň.

Zobrazenie archivovaných článkov

Webová aplikácia používateľovi ukazuje len názov článku, spomínané zločiny a ostatné metadáta. Ak chce vidieť jeho obsah, musí kliknúť na jeho názov, čo ho presunie na pôvodnú webstránku. Avšak môže nastať situácia, pri ktorej bude pôvodný článok z nejakého dôvodu odstránený. Preto chceme mať možnosť zobrazit' text článku aj v našej aplikácii, čo nie je problém, keďže ukladáme aj samotné HTML. Pri nájdených výsledkoch bolo pridané tlačidlo, ktoré používateľovi prehľadne zobrazí daný článok, aj s jeho pôvodným formátovaním.

PDF report

Jedným z hlavných cieľov nášho produktu, bola možnosť generovania PDF reportu. Avšak táto funkcionality bude dostupná len pre registrovaných používateľov. Tí budú mať pri prezeraní nájdených výsledkov možnosť pridať/odstrániť jednotlivé články do PDF reportu prostredníctvom jedného tlačidla. Jeho generovanie bude možné vykonať z novej obrazovky, kde okrem samotného článku a príslušných metadát vidíme, aký dopyt bol použitý pre jeho vyhľadanie.

Hotový produkt

Výsledkom globálnych cieľov za tento semester je funkčný produkt, ktorý poskytuje rozšírené možnosti vyhľadávania (filtrovanie podľa času, krajiny pôvodu a pod.) nad veľkým množstvom článkov. Ich počet stále stúpa, pretože proces zbierania bol automatizovaný a vykonáva sa každý deň. Produkt taktiež poskytuje aj možnosť registrácie a prihlásenia, čo používateľom odomyká novú funkcionality v podobe generovania PDF reportu. Napriek rozšírenej funkcionality si webová aplikácia zachovala svoj jednoduchý vzhľad.

Celkový pohľad na systém

V tejto kapitole sú zobrazené diagramy, ktorými sme sa riadili počas implementácie tohto systému alebo vznikli neskôr na zlepšenie komunikácie v tíme. Samotné diagramy približujú fungovanie aktuálnej verzie systému Adverse Media Screening.

Vysoká architektúra

Na obrázku nižšie je zobrazená vysoká architektúra s modulmi, nástrojmi a databázami, ktoré aktuálne využívame.

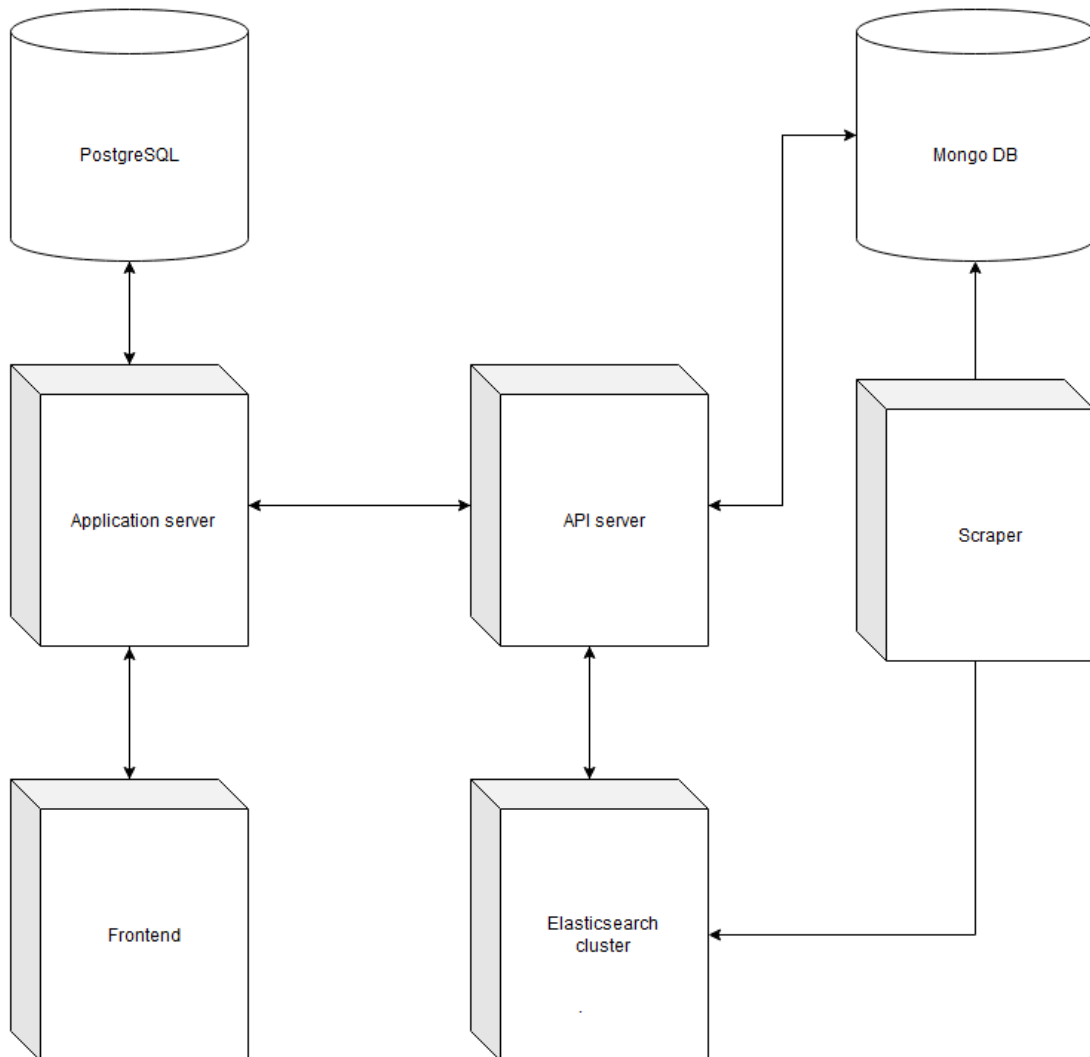
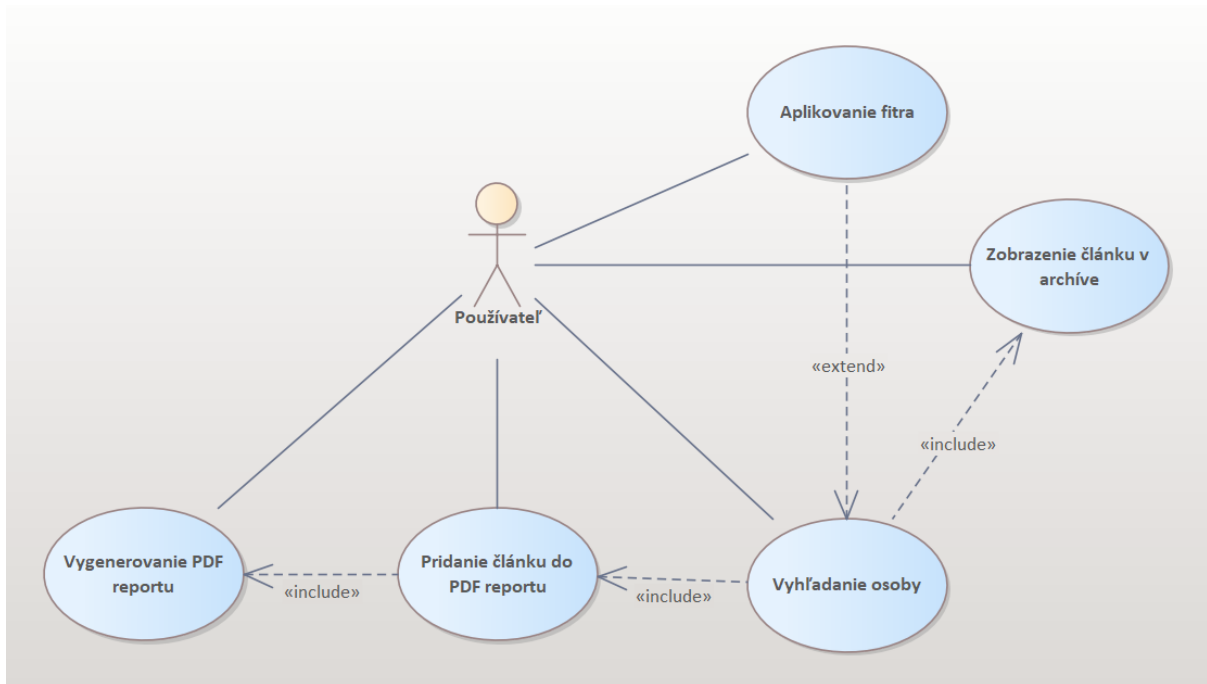
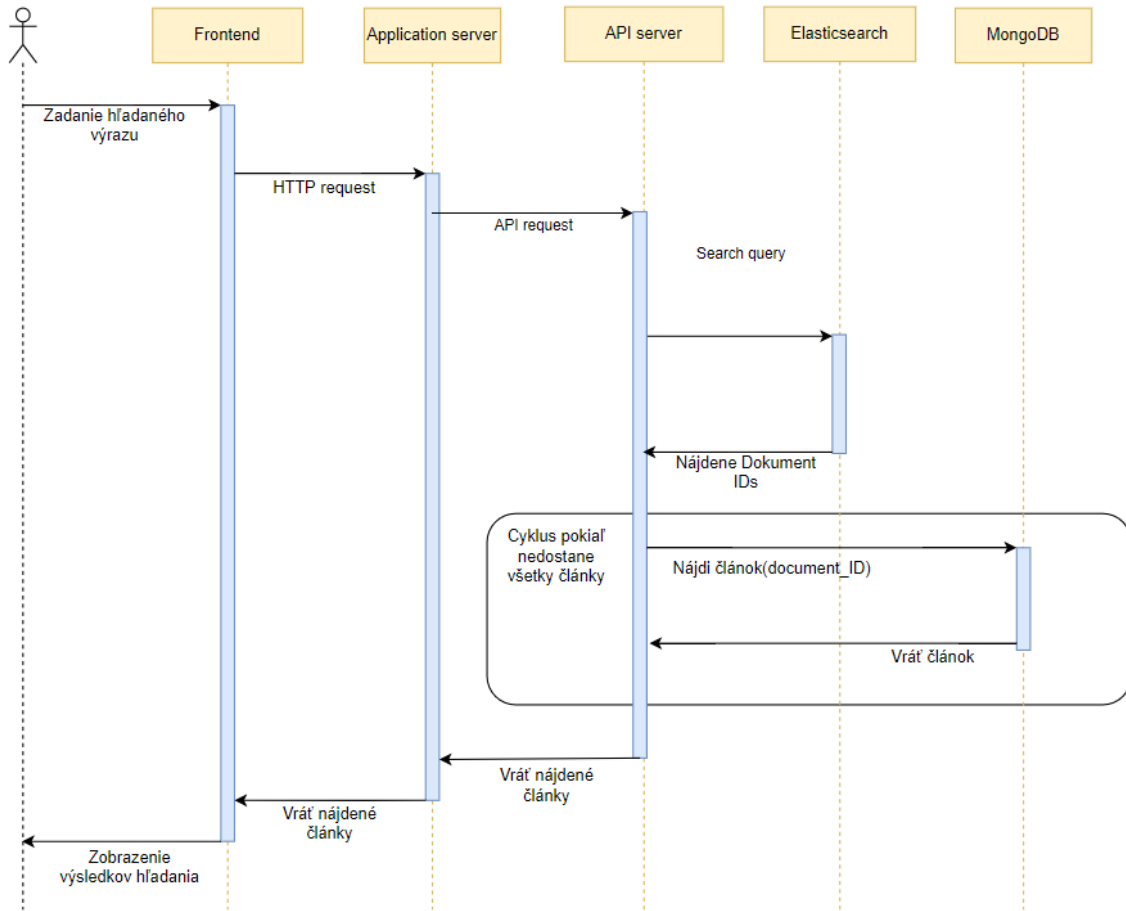


Diagram prípadov použitia

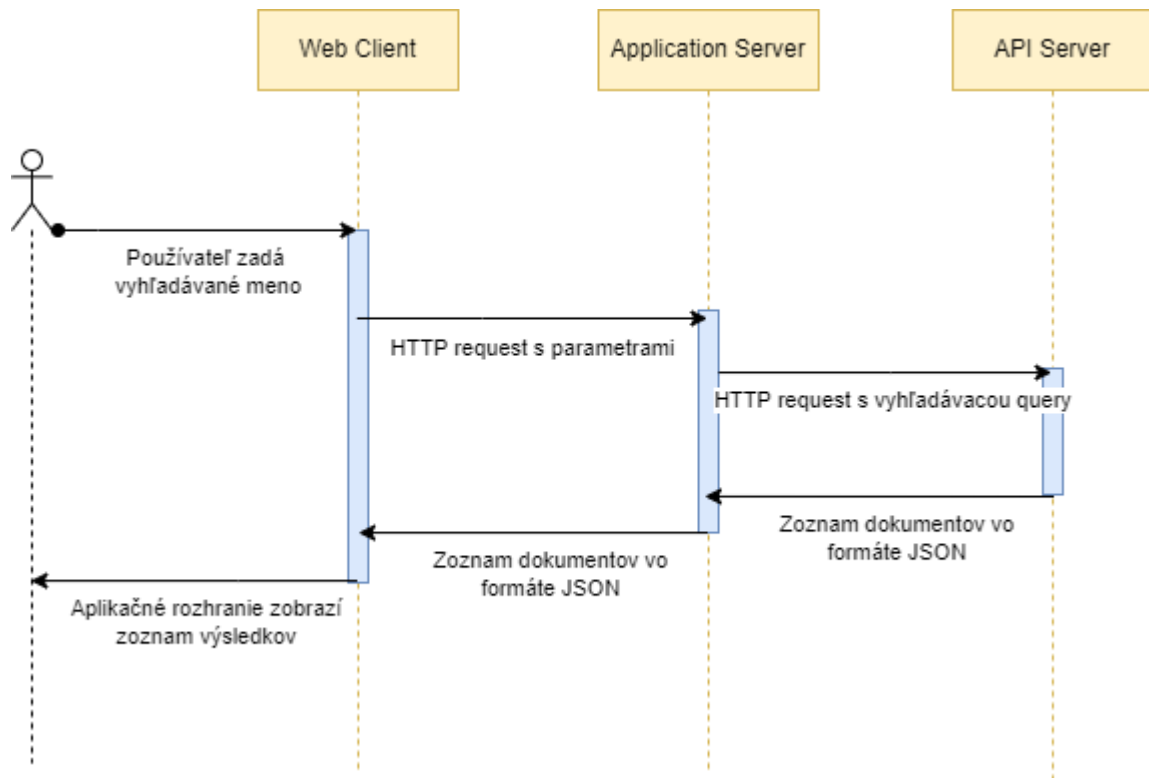
Na obrázku nižšie je zobrazený diagram prípadov použitia, ktoré je aktuálne používateľ prostredníctvom nášho systému schopný vykonávať.



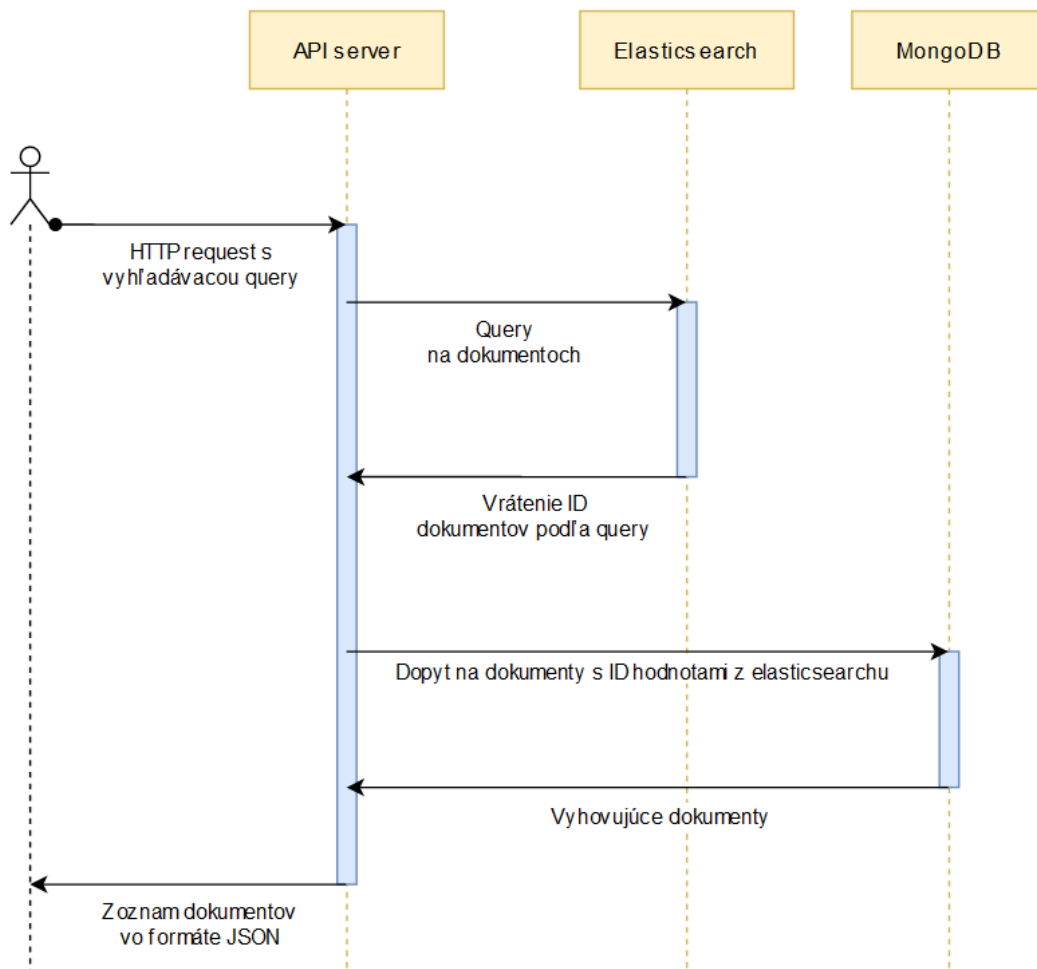
Sekvenčný diagram vyhľadávania v systéme



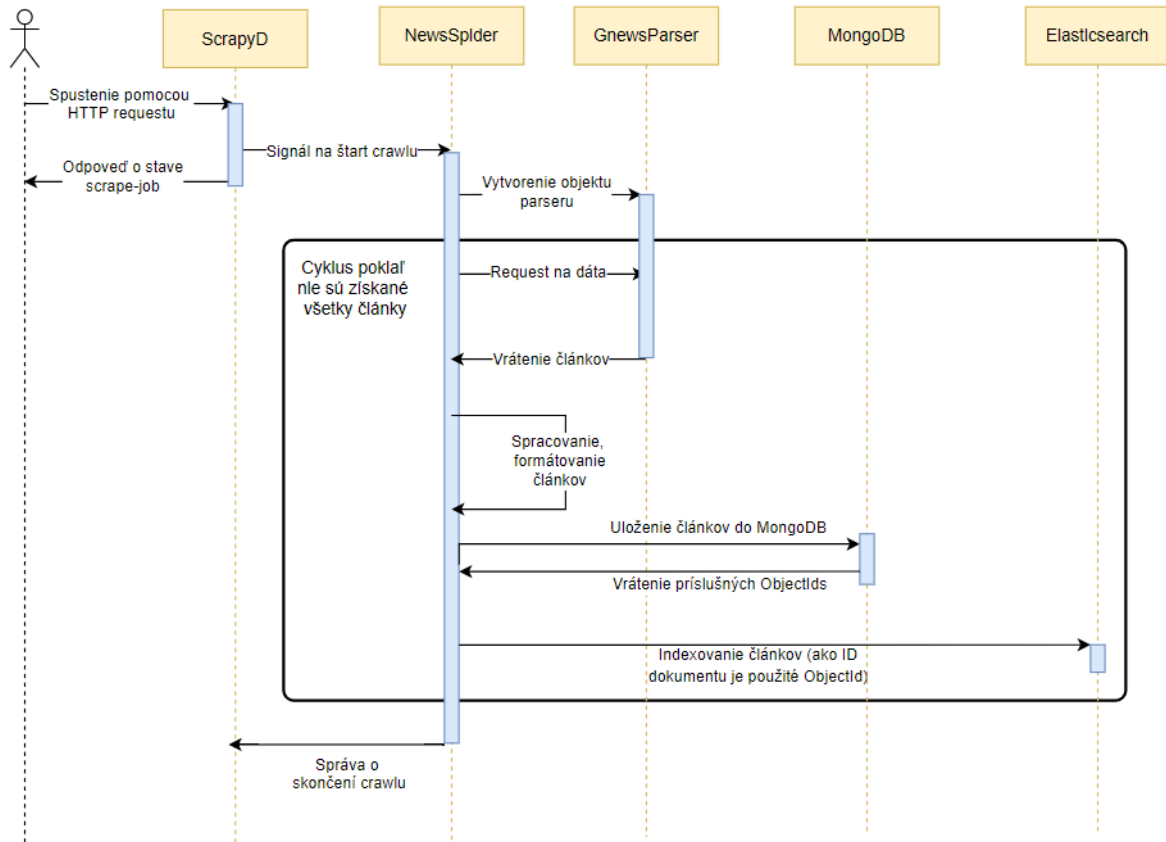
Sekvenčný diagram fungovania webovej aplikácie



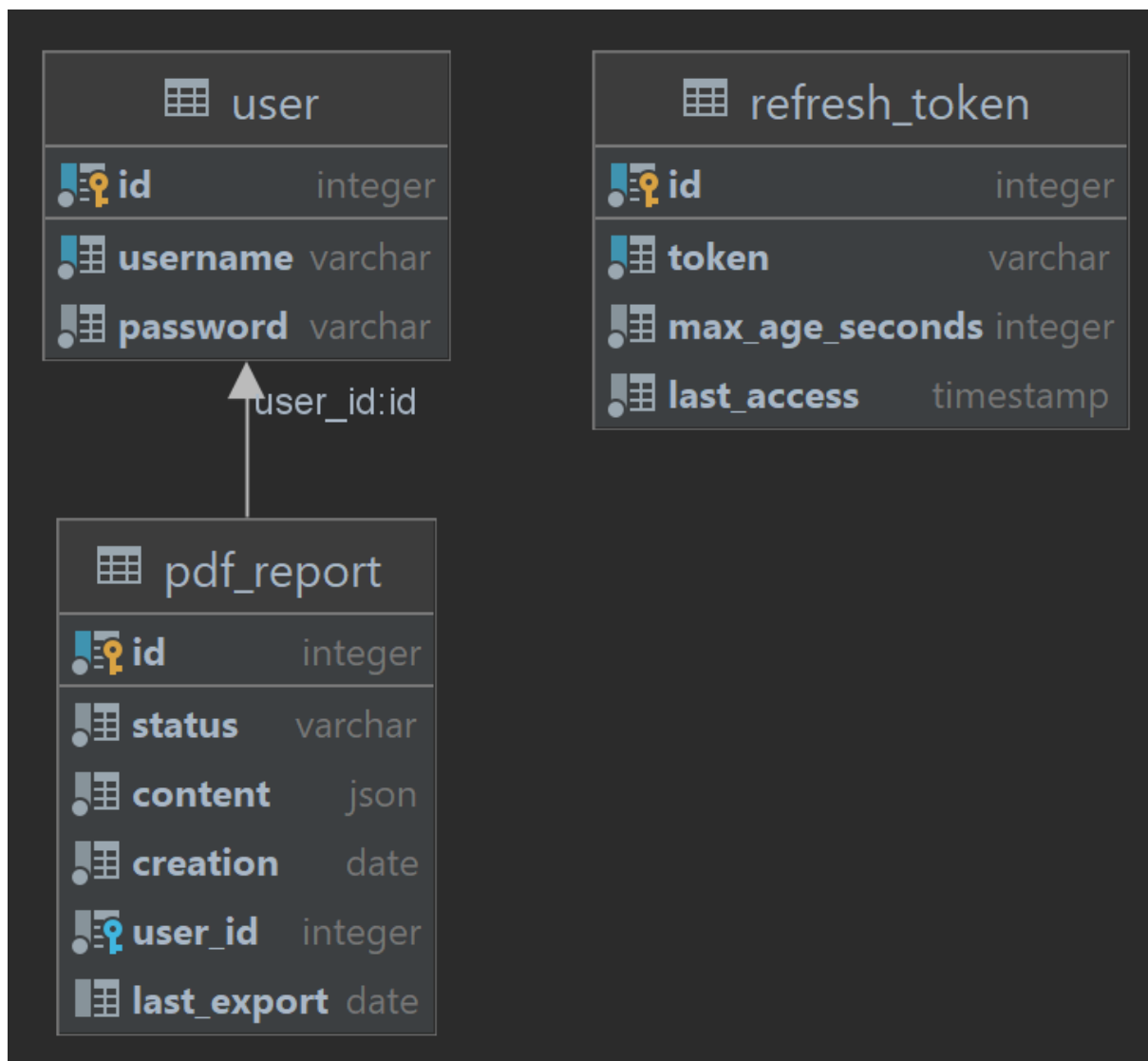
Sekvenčný diagram fungovania API serveru



Sekvenčný diagram fungovania scrapovania



Dátový diagram PostgreSQL



Moduly systému

V kapitole sú uvedené jednotlivé moduly systému, ktoré sú opísané podľa ich analýzy, návrhu, implementácie a testovania. Jednotlivé moduly sú Elasticsearch, Indexer, Aplikačný server, Frontend, API server a Scrapy.

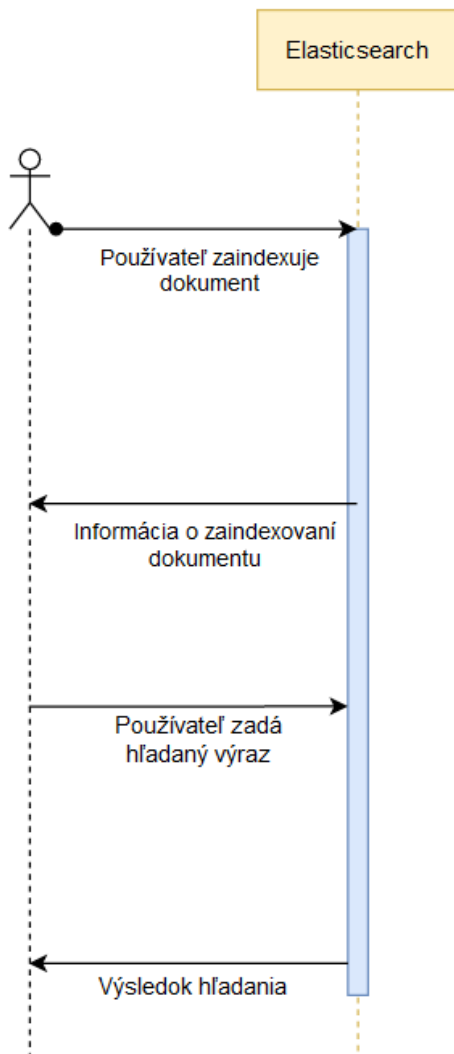
Modul Elasticsearch

Táto sekcia slúži na predstavenie modulu Elasticsearch, ktorý využívame na indexovanie obsahu stiahnutých článkov.

Analýza

Aby používateľ dokázal v systéme vyhľadávať články na základe hľadaného výrazu, je potrebné, aby boli obsahy článkov zaindexované.

Návrh



Implementácia

Elasticsearch je nasadený prostredníctvom technológie Docker. Aktuálne naše prostredie disponuje clustrom zloženom z 3 elasticsearch node. Nad samotným clustrom je taktiež spustená inštancia aplikácie Kibana, ktorú používame predovšetkým na debugging a monitoring stavu uzlov.

Elasticsearch nevyužíva pri indexovaní žiadny analyzátor pre stop-slová, keďže indexujeme články v rôznych jazykoch a elasticsearch neobsahuje analyzátor pre každý z nich. Tento kompromis nám však nijako neškodí, jediná nevýhoda je využitie väčšieho miesta na disku. Naše vyhľadávacie dopyty obsahujú mená, čiže konflikt so stop-slovami je málo pravdepodobný.

Každý dokument je zložený zo samotného textu článku a taktiež z metadát, ktoré článok obsahuje. Tieto metadáta obsahujú jeho zdroj, dátum zverejnenia a pod. Metadáta sú využité pri aplikovaní filtrov vyhľadávania.

Elasticsearch si hodnotu poľa neukladá, pretože je to v rámci našej funkcionality zbytočné a pri vyhľadávaní zaťažujúce.

Každý dokument má ID z Mongo databázy, vďaka čomu sa dokážeme z vyhľadávania v indexe jednoducho dostať k obsahu a metadátam článku, ktoré sú uložené v databáze.

Testovanie

Fungovanie modulu Elasticsearch bolo testované prostredníctvom jednotkových testov. V rámci modulu indexovania sa fungovanie Elasticsearch tiež testovalo prostredníctvom nástroja Kibana.

Modul Aplikačný server

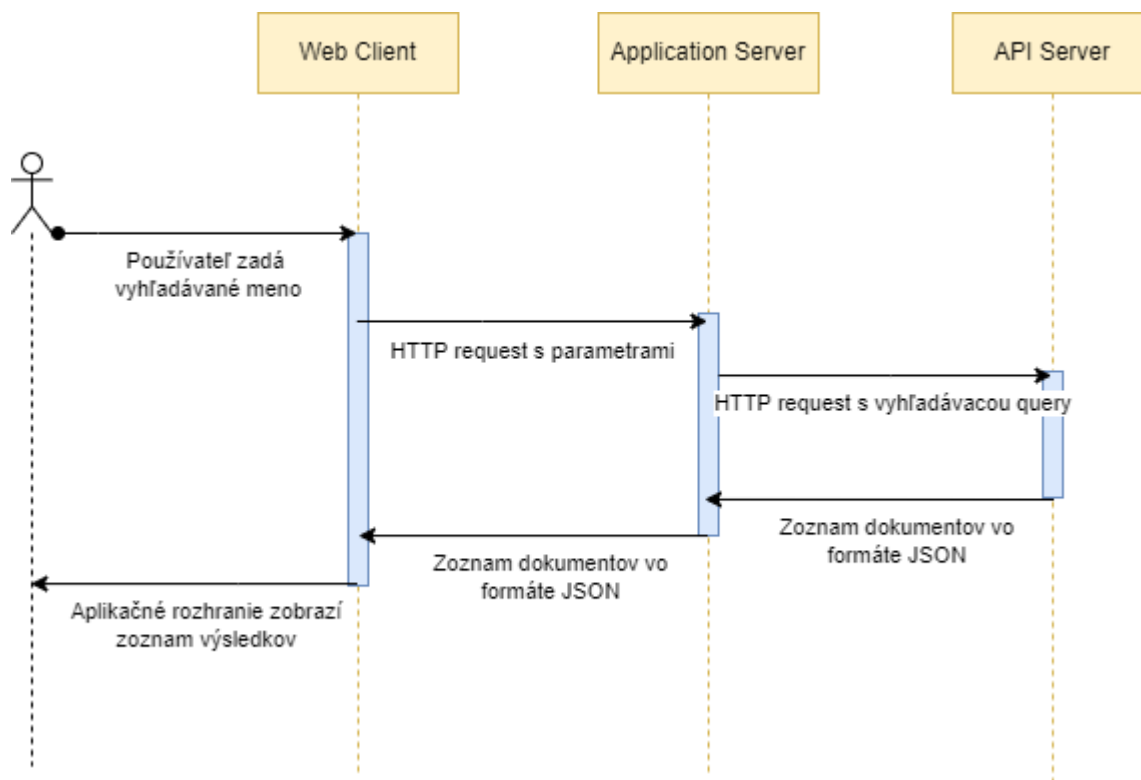
V tejto sekcii sa bližšie pozrieme na aplikačnú server.

Analýza

Aplikačný server zabezpečuje komunikáciu s ostatnými modulmi.

Návrh

V nasledujúcom sekvenčnom diagrame sa nachádza postupnosť krokov vedúca k zobrazeniu výsledkov vyhľadávania v prehliadači.



Implementácia

Implementácia je zabezpečená pomocou aplikačného servera, ktorý funguje pomocou knižnice ExpressJS. Jeho veľkou výhodou je dobrá organizácia do menších súborov pre napr. jednotlivé cesty, podcesty a databázové komunikátory

Webové rozhranie

Na strane servera sme implementovali webové API, ktoré sa aktuálne stará o podávanie statických súborov potrebných pre klienta. Taktiež sme zabezpečili preposielanie dát tak, aby Web API mohla spracovať žiadosť o dáta a preposlala ju na API server s využitím

Docker networkingu. Takéto preposielanie dát nám do budúcnosti zabezpečí jednoduchšiu implementáciu funkcionalít ako je napr. zaznamenávanie histórie.

Toto rozhranie sa ďalej stará o spracovanie prihlasovania a ukladanie používateľských údajov, ako je napr. rozpracovaný PDF report. Jednotlivé cesty sú bližšie dokumentované v našej živej dokumentácii.

Statické súbory

O požiadavky na statické súbory sa taktiež stará Express, čím držíme veľké množstvo konfigurácie na jednom mieste. Tým, že vybudovaná [React aplikácia](#) je statická, Express sa stará o odosielanie aj týchto súborov.

Aby sme mohli v Reacte používať React-Router, všetky požiadavky smerované na webový server, ktoré nie sú zachytené priamo definovanými cestami, vrátia vždy *index.js* React aplikácie. To znamená, že pri vyžiadaní */archive* a */about* sa vždy vráti ten istý súbor *index.js* a o render správnej podstránky sa stará frontendový React.

Prihlasovanie

Prihlasovanie je zabezpečené pomocou kombinácie JWT a session autentifikácie. Identifikačný json token je udržiavaný v dvoch cookie súboroch, kde jeden je dlhodobý a druhý krátkodobý. Tento token obsahuje identifikačné údaje používateľa, konkrétne meno a identifikačné číslo. Toto riešenie je zatiaľ komplikované kvôli aktuálne nedokončeného prechodu zo session based autentifikácie na JWT autentifikáciu.

Databáza

Ako databázu používame PostgreSQL. Pre zabezpečenie komunikácie používame knižnicu *pg*. Jednotlivé entity majú potom rozdelené komunikátory pre rôzne query rozdelené podľa logických súvislostí, kde napr. *user* a *pdf report* sú v rozdielnych súboroch.

Vývojové prostredie

Pri vývoji využívame kombináciu technológií. Vývoj backend a frontend časti aplikácie prebieha súčasne - full-stack prístup. Pre frontend využívame React development server (*npm start*) a pre backend využívame knižnicu *nodemon*. Takto sa nám pri vývoji vytvoria dva automaticky aktualizujúce sa servery miesto jedného. Toto riešime pomocou vlastného frontend API komunikátora, ktorý prepína cieľový server podľa stavu environmentálnych premenných. Toto riešenie nakoniec spájame s vlastným image databázy bežiacej v dockeri. Celkovo spúšťame všetky tri procesy spoločne pomocou knižnice *concurrently*, vďaka ktorej máme všetky tri výstupové streamy na jednom mieste oddelené vlastnými značkami.

Toto prostredie nám poskytuje rýchly prechod medzi staticky vybudovanou a dynamicky aktualizovanou verziou našej aplikácie bez akéhokoľvek prebytočného nastavovania.

Testovanie

Zhodnotenie funkcionality zatiaľ spočíva najmä z integračných testov, resp. kontroly komunikácie medzi aplikačnými modulmi (web client, application server) a vonkajšími modulmi (API server).

Tým, že naše cesty predstavujú prevažne jednoduché funkcie, jednotkové testovanie je zabezpečené pomocou nástrojov ako Postman, kde sa testujú rôzne vstupy pre jednotlivé cesty.

Modul Frontend

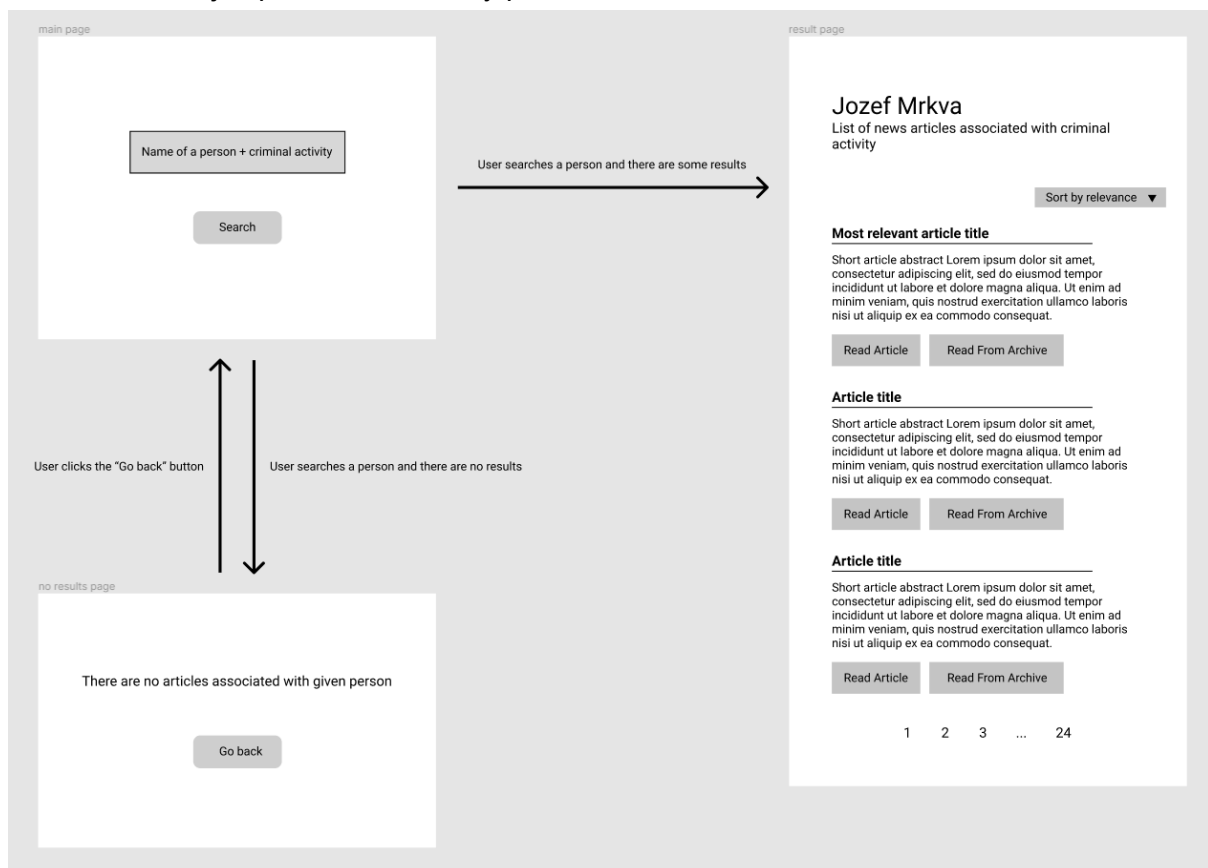
V tejto sekcii sa bližšie pozrieme na klientsku časť aplikácie - frontend.

Analýza

Frontend je potrebný pre zabezpečenie užívateľského rozhrania pre jednoduchší prístup na prácu s dátami. Webový klient slúži používateľovi ako grafické rozhranie.

Návrh

Návrh užívateľského rozhrania bol tvorený jednofázovo (Lo-Fi), kedy sme navrhli tri základné stavy, do ktorých sa používateľ môže dostať. Prvým je úvodná obrazovka, na ktorej používateľ zadá požadovaný reťazec, ktorý chce vyhľadať. Druhým je zobrazenie nájdených výsledkov a tretím je stav, kedy neboli nájdené žiadne výsledky. Nasledujúci obrázok obsahuje spomínané tri stavy užívateľského rozhrania.



Ostatné obrazovky boli implementované bez návrhu.

Implementácia

Implementácia je zabezpečená pomocou aplikácie napísanej v ReactJS a NodeJS.

Štýl dizajnu

Na jednoduchšie usporiadanie štýlu aplikácie používame MUI. MUI je knižnica užívateľského rozhrania React, ktorá poskytuje základné a pokročilé komponenty. Jednou z funkcií MUI je téma. Téma môže byť definovaná v základnom objekte s definovanými

požadovanými štýlmi a potom odovzdaná do funkcie `createTheme`. MUI poskytuje predvolenú sadu štýlov a časť z nich sme prepísali v súbore `Utils/Theme.js`. V paleta tém sme definovali dve hlavné farby pre celú aplikáciu, nazývané `primary` a `secondary`. Ostatné farby zostávajú nedotknuté a sú prístupné cez názov: `error`, `warning`, `info`, `success`. Výsledná paleta je definovaná takto:

Primary		
main #009688	light #b2dfdb	dark #004d40

Secondary		
main #757575	light #f5f5f5	dark #212121

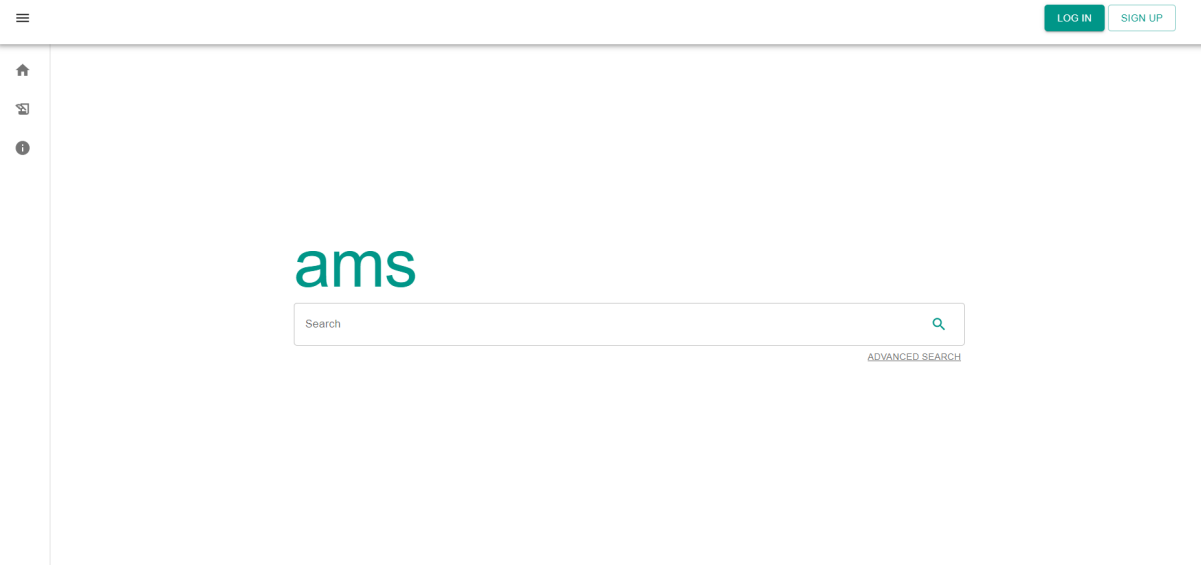
Domovská obrazovka

Východiskovou cestou (route) je `/search`. Obrazovka sa skladá z nasledovných komponentov:

- `TitleSearch` - komponent riadi vyhľadávanie a rozšírené vyhľadávanie.
- `MainLayout` - obsahuje nasledujúce komponenty:
 - `MainNavigation` - riadi otvorenie/zatvorenie bočného a horného panelu,
 - `TopBar` - horný panel obsahuje hlavný panel a riadi otvorenie/zatvorenie bočného panelu,
 - `MainBar` - hlavný panel obsahuje tlačidlo prihlásenia, resp. registrácie, ktoré otvorí dialóg pre prihlásenie, resp. registráciu,
 - `SideBar` - bočný panel riadi presmerovanie na nasledujúce podstránky:
 - domovská obrazovka - `/search`,
 - archív - `/archive`,
 - PDF report - `/pdf_report`,
 - stránka o projekte - `/about`,
 - `MainRouter` - riadi presmerovanie podľa nasledujúcej tabuľky:

Koncový bod	Komponent	Poznámka
<code>/</code>	<code><Navigate /></code>	Presmeruje na <code>/search</code>
<code>/search</code>	<code><TitleSearch /></code>	Domovská obrazovka
<code>/search/results</code>	<code><SearchResults /></code>	Výsledky vyhľadávania
<code>/archive</code>	<code><Archive /></code>	Archív
<code>/about</code>	<code><AboutPage /></code>	Stránka o projekte
<code>/pdf_report</code>	<code><ReportPDF /></code>	Správa PDF reportu

Dizajn domovskej obrazovky je zobrazený na nasledujúcom obrázku:



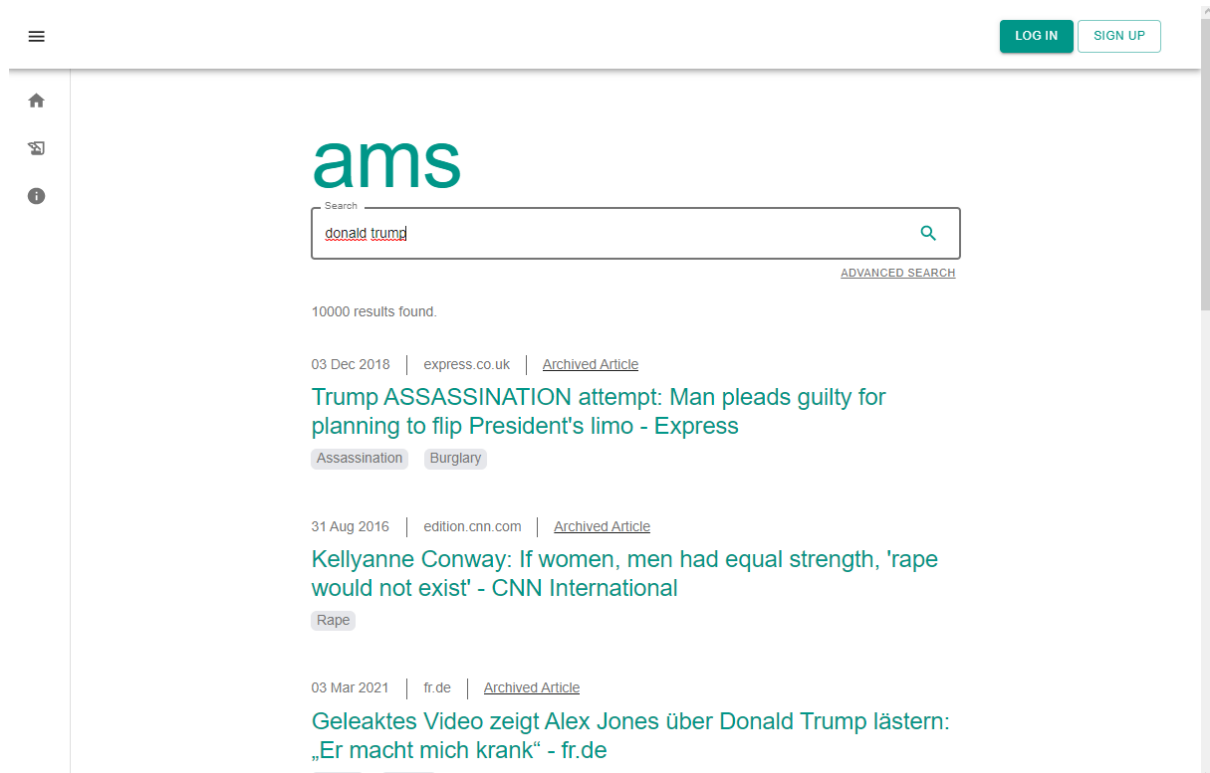
Základné vyhľadávanie

Základné vyhľadávanie je spravované dvoma komponentami:

- `TitleSearch` - po každom obnovení stránky prispôsobí jej stav podľa parametrov, ktoré sú uložené v URL. Parametre, ktoré ukladáme, sú:
 - `q` - aktuálna query, ktorá bola použitá na vyhľadávanie,
 - `page` - aktuálna strana nájdených výsledkov,
 - `from` - minimálny rok publikácie nájdených článkov,
 - `to` - maximálny rok publikácie nájdených článkov,
 - `regions` - regióny, kde boli nájdené články publikované,
 - `keywords` - kategórie kľúčových slov, ktoré sa v nájdených článkoch nachádzajú.Parametre `from`, `to`, `regions` a `keywords` slúžia pre správu rozšíreného vyhľadávania.
- `SearchResults` - zobrazenie výsledkov vyhľadávania.

Na domovskej stránke je možné zadať požadovaný reťazec, ktorý je poslaný na server. Odpoveď zo servera je následne zobrazená komponentom `SearchResults` v podobe zoznamu názvov, odkazov, dátumov publikovania a kľúčových slov pre nájdené články. Používateľa informujeme, koľko výsledkov bolo vrátených na požadovaný dopyt

Nasledujúci obrázok predstavuje zobrazenie výsledkov po zadaní dopytu "donald trump".



Rozšírené vyhľadávanie

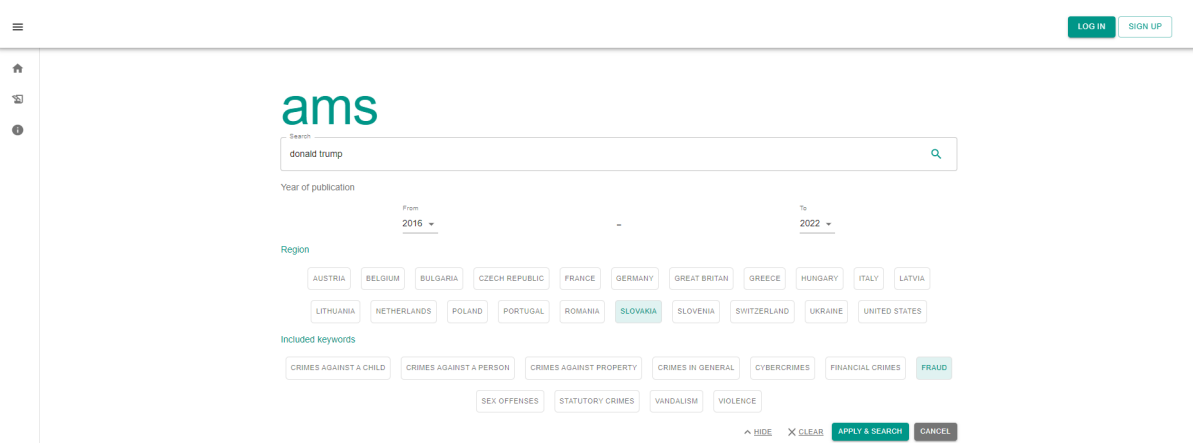
Filtre, ktoré v rámci rozšíreného vyhľadávania poskytujeme, sú:

- minimálny rok publikácie článkov (`yearFrom`)
- maximálny rok publikácie článkov (`yearTo`)
- región pôvodu článkov (`regions`)
- kategórie kľúčových slov, ktoré články obsahujú (`keywords`)

Rozšírené vyhľadávanie (implementované v komponente `AdvancedSearch`) je riadené komponentom `TitleSearch` pomocou nasledujúcich stavov:

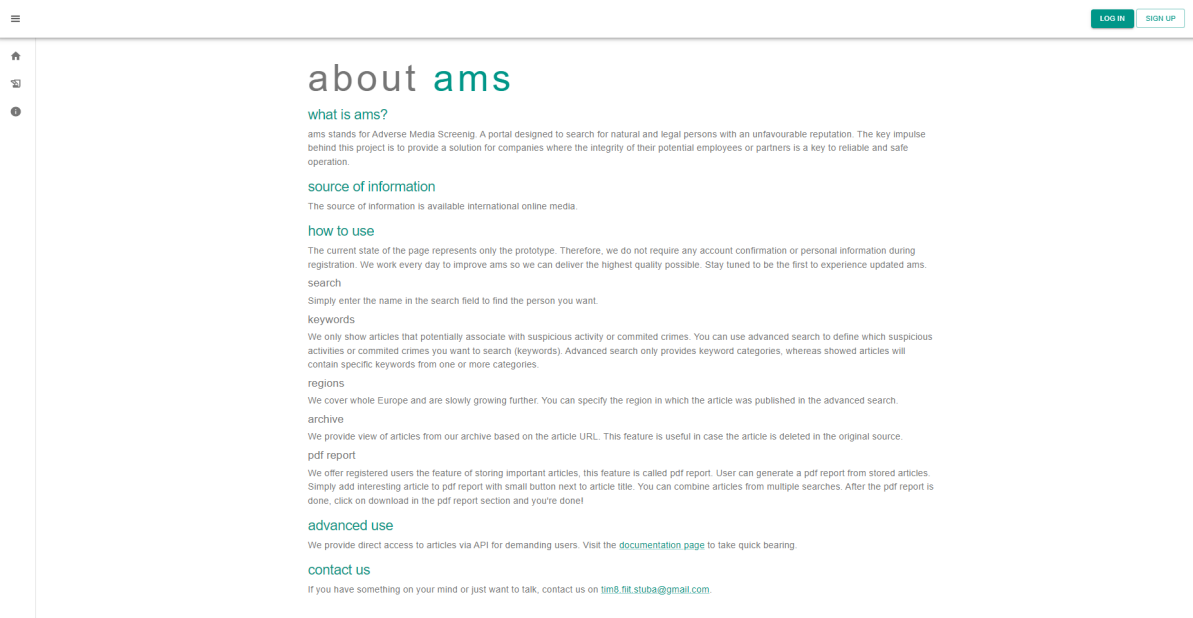
- zobrazenie,
- vybratie `yearFrom`,
- vybratie `yearTo`,
- vybratie `regions`,
- vybratie `keywords`,
- skrytie,
- vyčistenie obsahu,
- aplikovanie,
- zrušenie.

Na nasledujúcom obrázku je zobrazené rozšírené vyhľadávanie, so zvolenými filtrami `regions` a `keywords`.



Stránka o projekte

Keďže funkcionalita stránky sa rozrastá, pre používateľov sme poskytli stručný prehľad jednotlivých funkcionalít aplikácie. Prehľad je implementovaný v komponente `AboutPage`. Na nasledujúcom obrázku je zobrazená aktuálna verzia stránky o projekte:



Archív

Náš archív zabezpečuje, že pokiaľ je už cieľový článok v cieľovom médiu nedostupný, užívateľ sa i naďalej bude môcť dostať k aspoň časti jeho obsahu. Nedostupnosť článkov je veľmi bežná najmä kvôli periodickému premazávaniu starších článkov mnohých médií.

ams archive

Search URL

https://www.kenosha.com/news/local/crime-and-courts/pleasant-prairie-man-faces-fourth-owi-felony-drug-charge/article_fe8b3cc5-e9a9-5e6b-9cba-16b450ac7 🔍

Published 2021/11/16 in GB

[Pleasant Prairie man faces fourth OWI, felony drug charge - Kenosha News](#)

[Drug Possession](#) [Drug Trafficking](#) [Felony](#) [Murder](#)

Sign up for our Crime & Courts newsletter

Get the latest in local public safety news with this weekly email.

Pleasant Prairie man faces fourth OWI, felony drug charge

A traffic stop in Bristol last week led to the arrest of a 40-year-old Pleasant Prairie man on two felony charges.

Eric J. Schmitt, of the 8400 block of 66th Court, made his initial appearance in Kenosha County Circuit Court on Monday. Commissioner Larry Keating imposed a \$6,000 cash bond for Schmitt, who is due back next week Tuesday for a 10 a.m. preliminary hearing.

Schmitt is charged with felony counts of possession with intent to deliver cocaine, which carries a maximum fine of \$100,000 and 25 years in prison, along with his fourth drunken-driving offense, a felony that carries a possible fine of \$10,000 and six years in prison.

According to the criminal complaint, a deputy with the Kenosha County Sheriff's Department on patrol on 79th Street at Highway 45 observed the defendant's vehicle with a suspended registration.

When the deputy spoke to Schmitt, he reported noticing a strong odor of intoxicants. Schmitt admitted to having one drink after work before he drove home, the complaint states.

PDF report

Táto funkcionálnosť je prístupná iba prihlásenému užívateľovi. Ak počas vyhľadávania informácií o osobe narazí na viacero dôležitých článkov, môže ich pridať do svojho reportu. Akonáhle používateľ ukončil screening a chce vyhľadané články s niekým zdieľať, môže tak vykonať prostredníctvom PDF súboru. V tomto súbore budú obsiahnuté všetky články z reportu, pričom každý z nich bude mať zvýraznené riadky textu, v ktorých sa nachádza termín, podľa ktorého používateľ tento článok našiel (viď obrázok nižšie).

Trump ASSASSINATION attempt: Man pleads guilty for planning to flip President's limo - Express

Keywords: Assassination, Burglary

Found by term: donald trump

[express.co.uk](https://www.express.co.uk)

Express. Home of the Daily and Sunday Express. Trump ASSASSINATION attempt: Man pleads guilty for planning to flip President's limoA MAN pleaded guilty in a federal court on Friday to stealing a forklift truck with the intent to "flip" Donald Trump's limousine and kill him. We use your sign-up to provide content in ways you've consented to and to improve our understanding of you. This may include adverts from us and 3rd parties based on our understanding. You can unsubscribe at any time. More infoGregory Lee Leingang, 42, was arrested in 2017 for stealing a forklift from a restricted building and attempting to enter Trump's motorcade route. At the time, the US President was visiting Mandan, North Dakota to speak on a tax reform at an oil refinery. US Assistant State Attorney Brandi Sasse Russell said Leingang wanted to "kill the President".Related articlesHe said: "The intent was basically try to get the limo, flip the limo and get to the President and he wanted to kill the President."Leingang pleaded guilty to separate charges of burglary and starting fires at the Bismarck Municipal Ballpark maintenance shop and the state parole and probation office.He was sentenced to 10 years in prison for the fires.He received two further five year sentences for stealing the forklift and for another burglary incident. His release date is expected to be in 2038.A sentencing hearing for Leingang is scheduled for February 15.Leingang's public defender said he was "suffering a serious psychiatric crisis" during the incident.According to local media, he is seeing a psychiatrist and therapist in prison.Related articles It is not the first time someone has attempted to kill the American firebrand.In 2016, Michael Steven Sandford was arrested at a Donald Trump presidential campaign rally in Las Vegas, Nevada after he attempted to seize the pistol of a police officer providing security for the event.Sandford, a 20-year-old British citizen who had a lengthy history of mental

Správa používateľa

Používateľ je spravovaný komponentom `UserProvider` pomocou React knižníc `createContext` a `useContext`.

Základný stav používateľa je definovaný objektom:

```
{
  username: {}
}
```

`UserProvider` poskytuje funkcie:

- prihlásenie,
- odhlásenie,
- registrácia,
- pridanie článku do PDF reportu,
- odobratie článku z PDF reportu,
- nastavenie stavu používateľa prostredníctvom cookie tokenu:
 - nastavenie stavu používateľa,
 - načítanie uložených článkov v PDF reporte.

Komponent je načítaný v základom komponente `App`, vďaka čomu je možné jeho funkcie využiť ktorýmkoľvek komponentom.

Testovanie

Frontend-ová časť aplikácie je testovaná manuálne. Po vrátení výsledkov zo servera sa výsledky vypíšu do konzoly a zároveň do používateľského rozhrania. Kontrolujeme, či sa výsledky zhodujú a či sa po zadaní nového dopytu korektne aktualizujú.

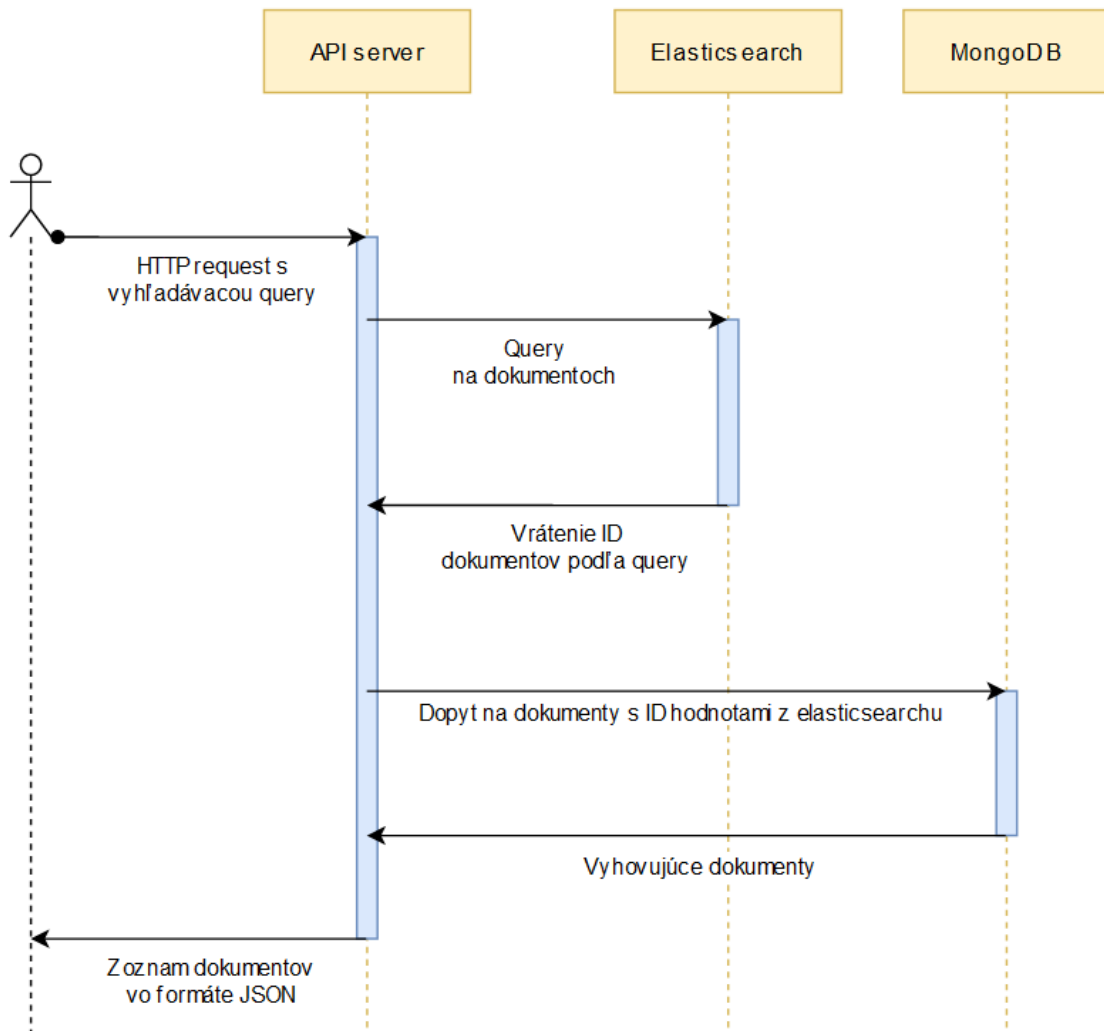
Modul API server

V tejto sekcii predstavíme API server nášho projektu.

Analýza

API server je potrebné implementovať pre oddelenie používateľského prostredia a hlavných biznis funkcií. API server spracováva vstupy od používateľského aplikačného serveru a vracia dáta do hlavného prostredia. Získava údaje z externých dátových zdrojov ako MongoDB a vyhľadáva v indexe dát prostredníctvom Elasticsearch. Komunikácia s týmto serverom je taktiež možná aj mimo aplikačného serveru, pre externé aplikácie a rýchle experimentovanie s dátami.

Návrh



Implementácia

API server je implementovaný pomocou jazyku Python a Flask frameworkom. Pomocou Flask triedy BluePrints sme implementovali verziovanie našej API. Pre vyhľadávanie bola vytvorená prvá API route `/search`. Tá akceptuje jeden povinný parameter `q`, ktorý obsahuje

vyhľadávaný reťazec. Ďalšie parametre sú smerované na redukciu vyhľadávania, ako napríklad časové rozpätie v rokoch, krajina pôvodu článku, kategórie hľadaných zločinov, číslo aktuálnej strany a počet článkov na stranu. Prvá verzia API však slúžila iba na testovanie.

Počas vývoja bolo implementovaných viacero verzií, no najaktuálnejšia je momentálne štvrtá. Všetky verzie od v2 sa dokážu pripojiť na modul Elasticsearch-u a podľa zadaného dopytu vyhľadať ID hodnoty dokumentov. Následne sa prostredníctvom získaných ID hodnôt z modulu MongoDB vyhľadajú vyhovujúce dokumenty, ktoré API server vráti webovej aplikácii vo formáte JSON. Okrem vyhľadávania boli vytvorené dve nové routes: */archive* a */report*. Prvá z nich dostáva ako parameter *link*, ktorý obsahuje url vybraného článku. Podľa tejto hodnoty vyberie a vráti požadovaný článok z databázy. Route */report* dostáva *ids* ako parameter, čo je zoznam ID hodnôt používateľom zvolených článkov. Tie vyberie z databázy a taktiež ich vráti.

Posledná route */stats* slúži na zobrazenie metadát o Mongo databáze a Elasticu. Zobrazí údaj o presnom počte dokumentov v oboch moduloch a taktiež aj ich veľkosť na disku.

Testovanie

API bolo testované manuálne, prostredníctvom HTTP requestov priamo na koncové body definované v zdrojovom kóde. Najjednoduchší spôsob tohoto testovania môže byť prostredníctvom jednoduchých command-line nástrojov ako curl alebo wget.

Taktiež bolo implementované jednotkové testovanie prostredníctvom vstavanej Python knižnice - *unittest*. Toto testovanie bolo zamerané na overovanie správneho fungovania API, pomocou knižnice *requests*.

Testovaná bola aj Mongo databáza. Rovnako, ako v predchádzajúcom prípade, aj tu boli uplatnené jednotkové testy s využitím knižnice *unittest*. Konkrétne sa jednalo o testovanie vyťahovania dokumentov z kolekcie *articles* a zisťovania, či boli vybrané správne.

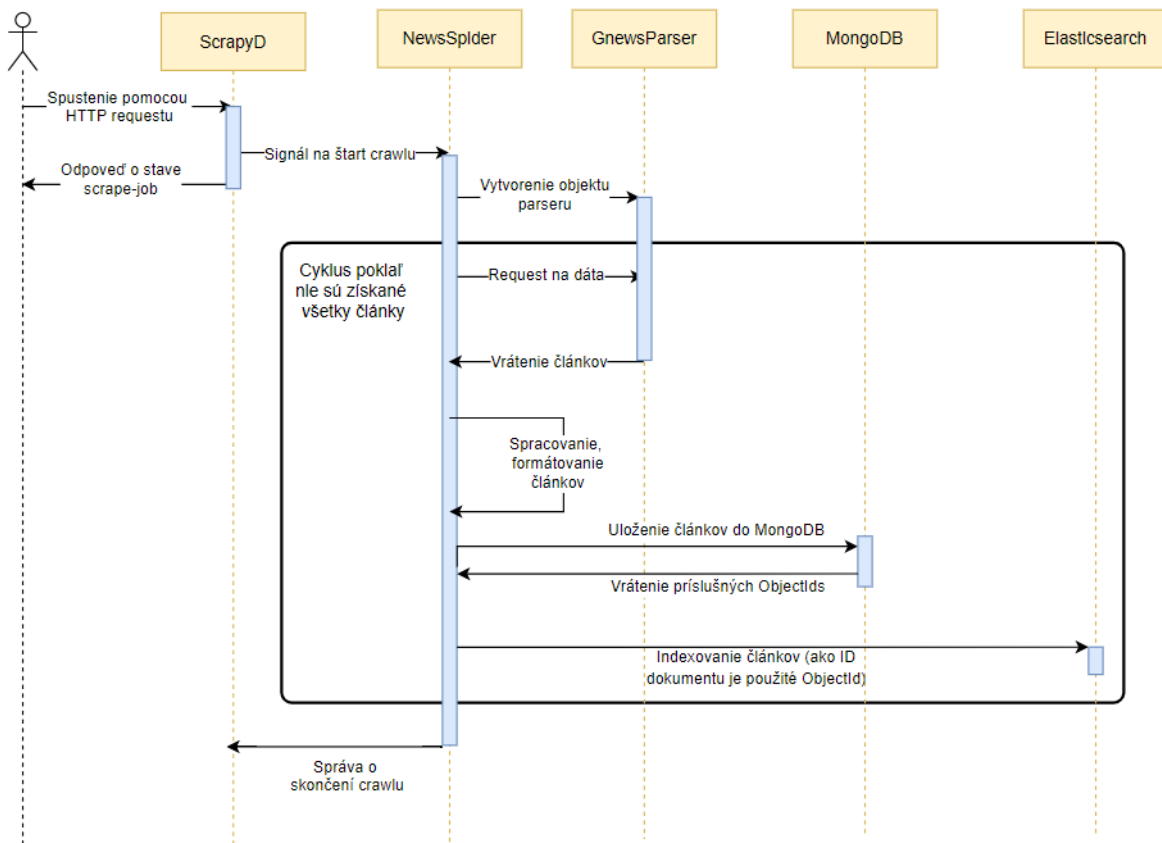
Modul Scraper

V tejto sekcii sa pozrieme na fungovanie modulu Scraper.

Analýza

Hlavnou úlohou scraperu bolo získať potrebné dáta z Google News - teda spravodajské články, ktoré nejako súvisia s trestnou činnosťou. Po získaní týchto článkov je potrebné ich vhodným spôsobom očistiť, aby zaberali čo najmenej miesta a zároveň nestratili svoju výpovednú hodnotu. Dáta sú následne uložené do Mongo databázy. Tá pre každý článok vráti ObjectId, ktoré použijeme ako ID dokumentu pri indexovaní v Elasticsearch-i.

Návrh



Implementácia

Scraper bol implementovaný v jazyku Python s využitím knižnice Scrapy. Ako zdroj článkov sme využili portál Google news a ich RSS stream API. Pre parsovanie tohto API bola vytvorená trieda GnewsParser, ktorá získa dátum zverejnenia, jazyk a odkaz na články z daného RSS feedu. Pomocou získaného odkazu sme cez Scrapy získali HTML kód článku a následne pomocou XPATH štandardu sme extrahovali všetky heading a paragraph tagy. Extrahovaný a redukovaný text sme uložili do MongoDB. V článkoch, kde sa nenachádzajú

tieto tagy, sme uložili celé HTML. Pomocou knižnice Scrapy sme dokázali tento proces paralelizovať prostredníctvom až 16 vlákien.

Na získavanie pre nás relevantných článkov (teda článkov, ktoré súvisia s nejakou trestnou činnosťou) sme použili ručne vytvorený zoznam trestných činov, ktorý obsahoval 96 názvov zločinov.

O získanom článku nakoniec ukladáme tieto informácie:

- názov článku,
- dátum uverejnenia článku,
- odkaz na daný článok,
- spracované telo článku,
- zločiny, pomocou ktorých sme tento článok získali,
- lokalizácia (krajina pôvodu + jazyk).

Scraper sa ovláda pomocou rozhrania scrapyD, čo je jednoduchá HTTP abstrakcia nad tzv. scrapy pavúkmi - programy smerované na spracovanie obsahu z konkrétnej domény. V našom projekte je definovaný len jeden pavúk - NewsSpider. Pomocou scrapyD dokážeme HTTP POST metódou odoslať príkaz na spustenie nášho pavúka spolu s parametrami na jeho ovládanie. Medzi tieto parametre patria:

- súbor zločinov, pomocou ktorých vyhľadávame články na Google News,
- počiatkový dátum vydania článkov,
- koncový dátum vydania článkov,
- lokalizácia (krajina pôvodu + jazyk).

Ďalšími HTTP metódami sa dá sledovať stav pavúka, zastaviť pavúka alebo čítať jeho logy.

Pomocou rozhrania scrapyD bolo implementované automatické získavanie článkov. Bol vytvorený tzv. plánovač získavania článkov, ktorý spúšťa náš modul scraperu pomocou HTTP POST request-ov každý deň o piatej hodine ráno, pričom získava spravodajské články z predošlého dňa. Takto dokážeme zaručiť aktuálnosť nami získaných článkov, keďže získame všetky články z predošlého dňa z každého nami dostupného regiónu.

Testovanie

Modul Scraper sa testoval manuálne. Bol spustený na menšej dátovej vzorke a jeho výstup sme ručne prešli. Spracované články boli porovnané s jeho HTML originálom najprv po obsahovej stránke a následne sa porovnalá redukcia vo veľkosti. Výpovedná hodnota článku ostala vo väčšine prípadov neporušená a veľkosť bola redukovaná v priemere o 93 %. Ďalšie testovanie bolo vykonávané pomocou jednotkového testovania. Pomocou vstavanej Python knižnice *unittest* bola testovaná časť programu zodpovedná za parsovanie Google news API. Kontroluje sa správnosť skladania URL pre dopyty na API, správnosť skladania query a taktiež krokovanie dní.

Prílohy

Protokol z testovania

Tester: Digital Systems a.s., Bnt international sro

Dátum odovzdania produktu na testovanie: 21.4.2022

Dátum prijatia protokolu: 17.5.2022

Hodnotenie funkcionality advanced search

Chyba mi tu hľadanie podľa jazyka, keďže angličtina môže byť aj vo viacerých regiónoch.

Chyba mi tu možnosť zoradenia výsledkov podľa dátumu.

Chýba mi to možnosť zadať si vlastný filter napríklad iné slovo, viacej špecifické. Napríklad špecifický podvod apod.

Pri hľadaní musím dať presne meno aj s diakritikou. Pri Jánovi Počiatkov som zadala Jan Počiatek a nič mi nenašlo. Tak ako aj preklep, ak dve písmenka som omylom prehodila, systém nič nevyhodil.

Google má nejaké intuitívne vyhľadávanie podľa lokácie, tuto mi ukazuje pri Vladimirovi Putinovi výsledky spreď mesiaca najskôr a ďalej z Rumunska a Talianska, čo je úplne irelevantné. Prosím porozmýšľať o nejakom defaultom prednastavení, keď si človek nič neprednastaví.

Hľadané meno ma viacero adverse media avšak vyhľadávač našiel iba jeden článok.

Ostatné chýbali. Niektoré mená nenašlo nič, avšak vieme, že v minulosti mali nejaké adverse media články.

Advanced search je možno len do roku 2016, ďalšia história chýba. Odporúčala by som aspoň 10 rokov nazad.

Hodnotenie funkcionality archive

Úprimne som nepochopila, kedy článok patrí do sekcie archív. Asi by som čakala, že Archive bude vyhľadávať na internete tam, kde to už nie je, napríklad prerobene = aktualizované internetové stránky apod.

Hodnotenie funkcionality pdf report

Exportovať možno len nájdené články jednotlivo, avšak export je dosť nečitateľný, urobila som printscreen a teda nie je príliš dobre použiteľný. Odporučila by som export do pdf s možnosťou dopísať tam výsledok posúdenia a ohodnotenia vyhľadávania s označením kto komentár vytvára a kedy.

Nenašla som export vyhľadávacích výsledkov do pdf. Túto funkcionality som tu nenašla, tento report je možno len ako print to pdf. Teda je možno exportovať do pdf len prvú stranu. Ideálne je export celého hľadania s možnosťou dopísať tam výsledok posúdenia a ohodnotenia vyhľadávania s označením kto komentár vytvára a kedy.

Hodnotenie celkovej práce s prototypom

(napr. Bolo používanie jednotlivých funkcionalít intuitívne?)

Pekné a sympatické, avšak vysvetlenie skratky AMS by som odporúčala hneď na prvej stránke, keďže mi to nie je jasné. AMS je všeobecne skratka pre Amsterdam. Aby bolo jasne na čo je ten search zameraný.

Veľmi pekný design avšak intuitívne mi trochu dlhšie trvalo, keď som pochopila na čo je pdf report. Tiež som nepochopila úplne princíp toho archived search, keďže článok, ktorý som vyhľadala sa mi ukázal, keď som naň klikla a nemusela som úplne použiť túto funkciu.

Ďalšie postrehy

Tiež úprimne, ako zákazník by som asi nemala príliš veľa dôvodov používať tento tool namiesto google, keďže tej je flexibilnejší a export je asi podobný.

Technická dokumentácia

Všetky docker containery bežia na Docker verzii 20.10.10

Python programy fungujú na verzii 3.9

NodeJS verzia: 14

ExpressJS verzia: 4.7.1

ReactJS verzia: 17.0.1

Link na organizáciu so všetkými repozitármi: [FIIT-TEAM8](#).

Dokumentácia, ktorú sme využívali počas semestra, je možné nájsť na [linku](#).

Dokumentácia našej API je prístupná na portáli [SwaggerHub](#).

Riadenie projektu

Dokumentácia

Členovia tímu: Jakub Hlavačka, Dominik Horváth, Jakub Müller, Táňa Poláková, Dávid Silady, Adam Šípka

Vedúci tímu: Ing. Richard Marko, PhD.

Študijný program: Inteligentné softvérové systémy

Ročník: 1.

Kontakt na tím: tim8.fiit.stuba@gmail.com

Big Picture	4
Úvod	4
Práca v tíme	5
Jakub Hlavačka	6
Dominik Horváth	6
Jakub Müller	6
Táňa Poláková	6
Dávid Silady	7
Adam Šípka	7
Manažment tímu	8
Každý člen tímu je zodpovedný za svoju rolu	8
Na riadenie projektu využívame metodiku Scrum	8
Sme tímoví hráči	9
Šprinty	10
Šprint č. 1	12
Šprint č. 2	14
Šprint č. 3	16
Šprint č. 4	18
Šprint č. 5	20
Šprint č. 6	22
Šprint č. 7	24
Šprint č. 8	26
Šprint č. 9	28
Šprint č. 10	30
Globálna retrospektíva	33
Zimný semester	33
Letný semester	33
Motivačný dokument	35
Tím	35
Motivácia	35
Téma č. 4 - Adverse Media Screening [AMS]	36
Téma č. 8 - Educational Content Engineering Hub - Databáza otázok, odpovedí, úloh a riešení [ECEH-DU]	37
Príloha A - Zoradenie všetkých tém podľa priority	38
Príloha B - Rozvrh tímu	39
Metodiky	40
Metodika Scrum	41
Standup	41
Šprint review	41
Plánovanie	41
Práca v nástroji Jira	41
Backlog	42
Nástenka	42

Riadenie projektu

Git Metodika	42
Vetvenie	43
Commit	43
Pull requests	43
Metodika nasadzovania	43
Docker	44
Komunikácia kontajnerov	44
GitHub Actions	44
Build & Deploy	44
Definition of ready	45
Definition of done	46
Export evidencie úloh	46

Big Picture

Úvod

Riadenie projektu je kľúčový aspekt, ktorý ovplyvňuje či bude projekt úspešne ukončený, alebo nie. Zo začiatku zimného semestra sme tejto oblasti venovali málo pozornosti, čo sa čiastočne odzrkadlilo aj na našich priebežných výsledkoch. Človek sa najlepšie učí na vlastných chybách a my sme si dali záväzok neodsúvať náležitosti agilného vývoja bokom. Bolo však nevyhnutné pristúpiť k úpravám metodiky Scrum, ktoré nám zaručili relatívne hladký priebeh letného semestra. Prispôsobenie Scrum-u je celkom bežná prax, avšak školská pôda nám priniesla nové nečakané výzvy.

Počas jednotlivých šprintov sme sa neustále točili okolo zle odhadnutých bodov za príbehy. Nedokázali sme nájsť bodový kompromis medzi náročným príbehom, na ktorom ale nie je potrebné stráviť veľa času a nad nenáročným, ktorý si vyžaduje dlhé hodiny manuálnej práce. Na rozdiel od praxe, naša “pracovná doba” musela byť rozdelená medzi x predmetov, čo malo za následok mapovanie bodov za príbehy na odhad hodín, ktoré na danom príbehu strávime. Sme si vedomí, že tento prístup je zastaralý, ale naše výsledky sú dôkazom toho, že nám to takto fungovalo.

Aby tím dbal na každú oblasť či už implementačnú, dokumentačnú alebo inú, v súvislosti s vývojom softvéru, určili sme role pre každého člena tímu. Role členov tímu, ako aj ich plnenie a manažment úloh sme zdokumentovali v kapitole Práca v tíme. Aby nám v tíme fungovalo všetko na 100 %, boli potrebné určité manažérske schopnosti a procesy, ktoré sme definovali v kapitole Manažment tímu. Prvky agilného prístupu – šprinty, k vývoju inžinierskeho diela, sme zdokumentovali v kapitole Šprinty. V poslednej kapitole sme opísali globálnu retrospektívu riadenia projektu za celý akademický rok.

Práca v tíme

Spočiatku sme fungovali na princípe, že všetci robia všetko. Domnievali sme sa, že ak bude každý z nás zahrnutý do všetkých prebiehajúcich procesov či už implementačných, alebo nie, bude to pre nás veľké plus. S pribúdajúcimi týždňami zimného semestra sme si uvedomili, že týmto prístupom sme dospeli k stavu, kedy nikto nie je zodpovedný za nič. Náš posun vpred to výrazne spomalilo. Pochopili sme, že ak človek dostane svoju rolu v rámci tímu, môže sa zdokonaľovať iba v jednej oblasti namiesto všetkých.

Nasledujúca tabuľka je zhrnutím odsekov nižšie pre rýchlejšiu orientáciu.

Meno	Rola	Hlavné zameranie	Autorstvo dokumentácie
Jakub Hlavačka	Manažér dokumentácie inžinierskeho diela	Webová aplikácia, Dokumentácia	Big Picture - inžinierske dielo, Sekvenčný diagram vyhľadávania v systéme, Modul - frontend
Dominik Horváth	Manažér plánovania	Dáta, API server	Diagram prípadu použitia, Modul - API server, Modul - scraper
Jakub Müller	Code review, Manažér kvality	Stránka tímu, Dáta, Testovanie	Modul - API server, Modul - scraper
Táňa Poláková	Scrum master, Manažér komunikácie, Manažér dokumentácie riadenia projektu	Jira, Frontend webovej aplikácie, Dokumentácia	Big Picture - riadenie projektu, Export úloh, Modul - frontend
Dávid Silady	Hlavný architekt	Webová aplikácia	Modul - aplikačný server, frontend
Adam Šípka	Manažér testovania	Dáta, Databáza článkov	Dokumentácia metodík, Ciele za letný semester

Snažíme sa, aby nikto z nás nemal pocit, že niekto robí viac, prípadne menej. Našu snahu, samozrejme, podkladáme počtom hodín, ktoré za každý šprint odrobíme. Celkový podiel práce sa nachádza v tabuľke nižšie.

Meno	Podiel práce (%)
Jakub Hlavačka	16.66667
Dominik Horváth	16.66667

Jakub Müller	16.66667
Táňa Poláková	16.66667
Dávid Silady	16.66667
Adam Šípka	16.66667

V ďalších odstavcoch vám predstavíme prácu jednotlivých členov tímu o niečo podrobnejšie.

Jakub Hlavačka

Na začiatku tímového projektu sa Jakub venoval úlohám spojeným s dátami a vývojovým prostredím. Neskôr sa však zameril na webovú aplikáciu, konkrétne funkcionálne spojenie s PDF reportom. Z manažérskej perspektívy sme sa ako tím zhodli, že je vhodný adept na dokumentovanie inžinierskeho diela. Samozrejme, jednotlivé časti softvéru budú spravidla dokumentovať práve ich vývojári, avšak Jakub dá dokumentu základnú štruktúru a bude sa oň starať. Jakub zdokumentoval Big Picture inžinierskeho diela a časť frontendu.

Dominik Horváth

Dominik venoval väčšinu svojho času vývoju modulov scraper a Flask API. Taktiež sa podieľal na úlohách spadajúcich pod DevOps charakteristiky, ako napríklad manažment hesiel v tíme, rozšírenie ElasticSearch clustra na 3 uzly, konfigurácia SSL komunikácie medzi uzlami, automatické pipeline a podobne. V kontexte zbierania dát vytvoril jednoduchý plánovač automatického každodenného zbierania internetových článkov. V rámci dokumentácie sa podieľal na opísaní modulov ElasticSearch a Scraper.

Jakub Müller

Vďaka za našu peknú stránku patrí Jakubovi M., ktorý ju implementoval. Pomáhal aj pri mechanizme získania a čistenia dát. Jakub sa postaral aj o to, aby sme mali všetky kódy na jednom mieste na platforme GitHub. Aby sme zaručili kvalitu nášho kódu, Jakub sa podujal robiť tzv. code review a postavil sa do role manažéra kvality. Taktiež sa neskôr venoval aj implementácii testov, ktoré boli aj zautomatizované pomocou platformy Github Actions. V rámci dokumentácie sa venoval modulu Scraper a modulu API server.

Táňa Poláková

Tánina rola – manažér dokumentácie riadenia projektu, bola takmer od začiatku jasná, keď sa podujala písať denníky z našich stretnutí. Zhodli, že bude scrum master a postará sa o efektívne používanie nástroja na manažment úloh. Okrem formálnych a dokumentačných záležitostí sa taktiež podieľala na čistení dát a návrhu API. Nakoniec sa našla v tvorbe frontendu. V rámci dokumentácie sa venovala Big Picture riadenia projektu, exporty úloh upravila tak, aby bolo v nich ľahko vyhľadávateľné a taktiež písala a upravila zápisy zo stretnutí, ktoré sú na stránke tímu. V rámci modulov je spoluautorom dokumentácie modulu Fronted.

Dávid Silady

Dávidove skúsenosti s tvorbou webu sme zužitkovali pri našom vlastnom webe projektu. Svoj čas nevenoval takmer ničomu inému, než náležitosti súvisiacim s webom, ktorý svojim spôsobom zastrešoval. Medzi ne patrí aj návrh používateľského rozhrania, komunikácia medzi klientom a serverom, logika prihlasovania, ako aj globálna konfigurácia ciest našej domény. Dávid je spoluautorom dokumentácie vysokej architektúry a aj preto sme ho zvolili ako hlavného architekta. Dokumentácia modulu aplikačný server a časť frontendu je aj jeho zásluha.

Adam Šípka

V prvom semestri sa Adam primárne venoval Mongo databáze, ktorú spravuje dodnes. Taktiež vyriešil jej napojenie na náš scraper. Neskôr sa začal podieľať na vývoji Flask API, ktorú prepojil s Elasticom a so spomínanou Mongo databázou. Tu bolo potrebné napísať a následne dolaďovať dopyt, aby naša aplikácia vedela používateľom zobrazíť len relevantné výsledky. Rola, ktorej sa zaviazal, že sa jej bude poctivo venovať a študovať potrebné náležitosti, je manažér testovania. Okrem toho, že spolupracoval pri dokumentovaní modulu API servera, Adam taktiež zdokumentoval aj naše metodiky.

Manažment tímu

Ako sme už vyššie uviedli, nedá sa povedať, že v začiatkoch semestra by sme mali nejakú formu manažmentu tímu. V tejto chvíli už však fungujeme nasledovne:

Každý člen tímu je zodpovedný za svoju rolu

Čo to znamená? Napríklad, že scrum master sa postará o to, aby mal každý čo robiť. Je potrebné, aby kontroloval, či sa členovia tímu starajú o svoje príbehy v nástroji Jira. Určili sme si, že scrum master nebude stáť nad členmi tímu a kontrolovať, či niekto urobil menej, alebo viac. Za toto už berie každý osobnú zodpovednosť. Ak niekto vidí, že odrobil príbehy rýchlejšie, ako sme predpokladali, ohlásí to ostatným členom tímu a spoločne sa vymyslí ďalší príbeh.

Manažér plánovania je zodpovedný za to, aby sme nevybočovali s pôvodného plánu a aby sme boli schopní dosiahnuť malé ciele, ktoré sme si vopred definovali. Mal by sledovať, či jednotlivé príbehy vedú k naplneniu plánu, alebo nie.

Manažér kvality dohliada na každý vyprodukovaný kus kódu a snaží sa z autora vyťažiť čo najvyššiu kvalitu. Domnievame sa, že je to istou formou aj psychologický efekt. Človek, ktorý si je istý, že jeho kód nikto iný okrem neho nebude vidieť, ho s veľkou pravdepodobnosťou odflákne. Ak ale kód bude posudzovať niekto ďalší, pôvodný autor sa bude viac snažiť. Manažermi kvality sme boli všetci. Každý príbeh mal prideleného svojho manažéra kvality, ktorého sme určili na základe znalostí, ktoré daný človek mal.

Manažér komunikácie predchádza nedorozumeniam zapríčineným zlou komunikáciou. Ak je potrebné niečo odkomunikovať s vedúcim tímu na diaľku, postará sa o to.

Hlavný architekt má posledné slovo, pokiaľ sa jedná o architektúru systému. Je odborníkom vo svojej oblasti a pomáha ostatným členom tímu navrhnuť architektúru čo najlepšie.

Manažér testovania rozhoduje, ktoré časti systému sa dajú testovať. Odporúča, aké typy testov je vhodné implementovať a dohliadne, aby „doručené“ časti kódu boli vhodne otestované.

Manažéri dokumentácie sa postarajú o to, aby sa dokumentácia písala priebežne (!). Toto je naozaj oblasť, kde si chceme dať obzvlášť veľký pozor. Písať dokumentáciu na poslednú chvíľu je veľmi nepríjemné.

Na riadenie projektu využívame metodiku Scrum

Našou najnosnejšou metodikou je scrum. Táto metodika je formálne opísaná spolu s ďalšími v kapitole Metodiky. Scrum sme si pre potreby projektu prispôbili. Naše šprinty trvajú dva týždne, pričom sa snažíme pri plánovaní šprintu o to, aby každý člen odrobil približne 12 hodín. Je nám jasné, že budeme čeliť aj istým odchýlkam, avšak tieto rozdiely si navzájom akceptujeme. Body, ktoré jednotlivým príbehom prideliujeme, zodpovedajú náročnosti úlohy. V našom prípade sa náročnosť rovná počtu predpokladaných hodín, ktoré človek odpracuje. Ekvivalenciu bodov a hodín sme si prispôbili kvôli lepšiemu plánovaniu. Na konci šprintu každý člen tímu zodpovie na otázky, čo hodnotí na šprinte ako pozitívne a naopak, čo by

chcel zlepšiť. V rámci zhodnotenia šprintu taktiež uvádzame percentuálny podiel členov na výsledku šprintu. Podiel počítame ako celkový počet získaných bodov za príbehy a celkový počet bodov, ktoré tím mohol získať, a to je šesťnásobok čísla 13. Ak sa nám nepodarilo body na šprint prispôbiť číslu 13, za najvyšší možný počet bodov považujeme aktuálne najvyššie číslo bodov za príbeh. Zároveň, podiel je aplikovaný aj z pohľadu všetkých šprintov, a tak vieme zhodnotiť, či človek, ktorý počas jedného šprintu odrobil menej, dobehol svoj podiel v inom šprinte.

Sme tímoví hráči

Motivujeme a rešpektujeme sa. Veríme, že príjemný kolektív je základ dobre odvedenej práce. Ak niečo nefunguje, nezvaľujeme na seba vinu navzájom. Namiesto toho sa snažíme pochopiť, odkiaľ problém prišiel a nájsť spoločne riešenie, ktoré, v rámci možností, vyhovuje každému z nás.

Šprinty

Kapitola je venovaná dokumentácii prvých troch šprintov. Prvý šprint sme začali o týždeň neskôr oproti odporúčanému dátumu začiatku. Dôvodom bola naša nevedomosť, ktorá nás samozrejme neospravedlňuje. Prvý šprint začal 12. 10. 2021, čo je štvrtý týždeň semestra. Naša práca na projekte však začala už v druhom týždni, avšak tieto dva týždne neboli evidované ako oficiálny šprint, a preto úlohy, ktoré sme vtedy urobili, neprislúchajú žiadnemu šprintu (sú dostupné v exporte úloh). Fungovali sme v podstate v týždňových iteráciách bez zhodnotenia na konci iterácie.

Pri prvom šprinte sme sa oboznámili s pridelovaním bodov za jednotlivé úlohy/príbehy. Nedá sa však povedať, že by naše odhady zodpovedali realite, čo sa aj odrazilo pri šprint review. Taktiež sme si uvedomili, že ľudské zdroje neboli naplno využité a poučili sme sa, že do budúca určite naplánujeme viac úloh/príbehov. Prvý šprint je zdokumentovaný v sekcii Šprint č. 1.

Druhý šprint sa už javil profesionálnejšie. Odhad sa nám síce podarilo o trochu zlepšiť, ale stále je na čom pracovať. Šprint je zdokumentovaný v sekcii Šprint č. 2.

Bohužiaľ, podarilo sa nám do tretieho šprintu preniesť dva príbehy. Postihla nás neočakávaná situácia a deň pred ukončením šprintu nám pribudlo veľa nových úloh, ktoré súviseli s odovzdaním prvého kontrolného bodu. Poučili sme sa, že si veci nebudeme odkladať na poslednú chvíľu a že si určíme, kto sa bude priebežne o dokumentáciu starať. Tretí šprint je zdokumentovaný v sekcii Šprint č. 3.

Cieľom šprintu č. 4 bolo mať hotový prototyp. Prepájali sme jednotlivé moduly, takže naše pridelené príbehy navzájom od seba záviseli. Nakoniec sme všetko stihli a prototyp máme hotový. Do budúceho semestra sme si preniesli jeden príbeh, ktorý priamo s prototypom nesúvisel a jeho náročnosť sme neodhadli správne. Štvrtý šprint je zdokumentovaný v sekcii Šprint č. 4.

Posledný šprint zimného semestra sme venovali dokončeniu dokumentácie a prípraveniu dokumentov na odovzдание. Dopadol až prekvapivo dobre, čo na konci semestra veľmi oceňujeme. Piaty šprint je zdokumentovaný v sekcii Šprint č. 5.

Šprint po vianočných prázdninách sme sa rozhodli naplánovať s predstihom, aby sme, ako obvykle, plánovaním nezabli nárazovo vyše troch hodín. Faktor, s ktorým sme naďalej bojovali bol odhad bodov. Šiesty šprint je zdokumentovaný v sekcii Šprint č. 6.

Siedmy šprint sa niesol v pozitívnom duchu. Pasovali sme sa iba s prílišnými prepojeniami jednotlivých príbehov. Šprint je zdokumentovaný v sekcii Šprint č. 7.

Príbehy ôsmeho šprintu naberali na zaujímavosti. Konečne sme sa dostali k implementácii funkcionalít, ktoré nepatrili medzi základné minimum. Hodnotenie šprintu bolo aj vďaka tejto skutočnosti pozitívne. Množstvo implementácie na frontende nám ale prinieslo konflikty na GitHube. Ôsmy šprint je zdokumentovaný v sekcii Šprint č. 8.

Riadenie projektu

Cieľ deviateho šprintu bolo odovzdanie prototypu na testovanie, ktorý sa podarilo naplniť. V tomto šprinte sme neidentifikovali nedostatky. Deviaty šprint je zdokumentovaný v sekcii Šprint č. 8.

Posledný šprint letného semestra sme venovali písaniu dokumentácie ako aj tvoreniu ďalších náležitostí súvisiacich s TP CUP. Šprint sa z pôvodných dvoch týždňov predĺžil na štyri, keďže plánovanie ďalšieho šprintu na dokončenie dokumentácie nám prišiel ako zbytočný krok. Plynule sme pokračovali v (nových) úlohách aj v nasledujúcich dvoch týždňov. Opäť sme neidentifikovali nedostatky, aj keď sme ho boli nútení mierne preplánovať v dôsledku článku na IIT.SRC. Alokovali sme si aj čas na zapracovanie spätnej väzby k testovaniu, avšak k tomuto neprišlo. Desiaty šprint je zdokumentovaný v sekcii Šprint č. 10.

Šprint č. 1

Trvanie: 12. 10. 2021 – 26. 10. 2021

Čo hodnotíme pozitívne?

- Po zavedení šprintov máme väčší prehľad, kto čo urobí/il.

Čo chceme zlepšiť?

- Zistili sme, že máme veľa nevyužitých ľudských zdrojov. Určili sme si do budúca, že cez týždeň budeme venovať projektu viac hodín.
- Chceme sa snažiť rozdeliť úlohy tak, aby nemal niekto ťažšie a niekto ľahšie úlohy.
- Chceme sa naučiť pracovať s Jirou efektívnejšie.
- Volit' relevantnejšie príbehy, nie také „aby boli“.
- Kvôli nerelevantnosti ostala jedna úloha neukončená. Budeme sa jej venovať neskôr.
- Chceme zlepšiť svoj odhad pri pridelovaní počtu bodov.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Stránka tímu	Jakub Müller	6	áno	6
Otvorenie portov na serveri	Dávid Silady	6	áno	6
Vyriešiť "sudo su"	Táňa Poláková	2	áno	2
User DB	Táňa Poláková	6	nie	6
Prieskum DB	Jakub Müller	1	áno	1
Prihláška na TP CUP	Dávid Silady	2	áno	2
Set up Elasticsearch v docker container	Jakub Hlavačka	4	áno	4
Získanie dát pre prototyp	Jakub Hlavačka	4	áno	4
Získanie dát pre prototyp	Adam Šípka	8	áno	8
Získanie dát pre prototyp	Dominik Horváth	8	áno	8

Riadenie projektu

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	8	8	16.66667
Dominik Horváth	8	8	16.66667
Jakub Müller	8	8	16.66667
Táňa Poláková	8	8	16.66667
Dávid Silady	8	8	16.66667
Adam Šípka	8	8	16.66667

Šprint č. 2

Trvanie: 26. 10. 2021 – 09. 11. 2021

Čo hodnotíme pozitívne?

- Naplánovali sme si 62 hodín, čo vychádza, že každý z nás mal 1.6 hodiny rezervu. Nakoniec sme odrobili viac, a to 68 hodín, čo je oproti predošlým 43.5 posun vpred. Úlohy sme si pridali v priebehu šprintu.
- Náročnosti jednotlivých príbehov sa celkom vyrovnali.

Čo chceme zlepšiť?

- Chceme príbehy špecifikovať menej, aby sme príbehy nemuseli rušiť, ak narazíme na nejaký problém.
- Zistili sme, že náročnosť úlohy je výrazne odlišná, ak na úlohe pracuje človek, ktorý s danou technológiou má skúsenosti a človek, ktorý má minimálne skúsenosti.
- Ak príbeh pridáme človeku s minimálnymi skúsenosťami, chceme daný príbeh radšej rozdeliť na menšie.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Parsovanie tagov pomocou CSS selector	Jakub Müller	5	áno	13
Parsovanie tagov pomocou regex	Táňa Poláková	8	zrušený	3
Upratať Jiru	Táňa Poláková	2	áno	3
Analýza bezstratovej kompresie textu	Jakub Hlavačka	8	áno	8
Lokálna MongoDB so získanými dátami	Adam Šípka	8	áno	5
Request URL s odpoveďou inou ako 200 uložiť do súboru	Adam Šípka	3	áno	1
Pridať info o projekte a opísať členov tímu	Jakub Müller	3	áno	1
Rozšíriť docker compose	Dominik Horváth	13	áno	8
Zavesenie klienta	Dávid Silady	13	áno	13
Parsovanie tagov pomocou lxml, readability, trafiletura	Táňa Poláková	3	áno	8

Riadenie projektu

Pamäťová zložitosť Elasticsearch na vzorke dát	Jakub Hlavačka	5	nie	5
Rozšírenie google news scrapperu na lokáciu UK	Dominik Horváth	1	áno	1
Vyhľadanie RSS dvojčaťa k HTML článku - všeobecný postup	Dominik Horváth	8	áno	5
Napĺňanie MongoDB priamo pri scrapovaní	Adam Šípka	8	nie	8

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	16	13	16.66667
Dominik Horváth	22	13	16.66667
Jakub Müller	9	13	16.66667
Táňa Poláková	13	13	16.66667
Dávid Silady	13	13	16.66667
Adam Šípka	19	13	16.66667

Šprint č. 3

Trvanie: 09. 11. 2021 – 23. 11. 2021

Čo hodnotíme pozitívne?

- Máme rozšírené komponenty, majú viditeľnejšiu funkcionálnosť.
- Prelúskali sme už väčšinu technológií.
- Aklimatizovali sme sa na metodiku scrum.

Čo chceme zlepšiť?

- Chceme, aby sa jednotliví členovia tímu venovali viac svojim rolám.
- Chceme zabezpečiť, aby bol každý príbeh viac zdokumentovaný.
- Chceme, aby sa každý staral o svoj príbeh a v komentároch v nástroji Jira popísal, čo urobil. Taktiež do komentáru napíše komentár od vedúceho ako záznam zo stretnutia.
- Mali by sme uvažovať o pár týždňov dopredu pri plánovaní príbehov aby sa nestalo, že na poslednú chvíľu nám pribudne veľa nových – ako tomu bolo v tomto šprinte.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Napĺňanie MongoDB priamo pri scrapovaní	Adam Šípka	2	áno	3
Zabrániť zacykleným buildom	Dominik Horváth	5	áno	3
Rozšíriť úložisko pomocou nepriradeného disku	Dominik Horváth	5	áno	3
Integrácia scraperu a vylepšeného parseru	Jakub Müller	5	áno	8
Integrácia MongoDB a Elasticsearch	Jakub Hlavačka	8	áno	8
Úprava dokumentácie zo stretnutí a šprintov tak, aby mohli ísť na stránku tímu	Táňa Poláková	8	áno	8
Zabezpečenie komunikácie medzi klientom a serverom	David Silady	8	áno	8
Analýza možností implementácie testovania	Adam Šípka	5	áno	5

Riadenie projektu

Návrh Flask API	Táňa Poláková	5	áno	5
Implementácia dummy Flask API	Dominik Horváth	5	áno	5
Rozšírenie docker compose o MongoDB a Elasticsearch	Jakub Hlavačka	5	áno	3
Dokumentácia k inžinierskemu dielu	Jakub Hlavačka	5	áno	5
Dokumentácia k inžinierskemu dielu	Dominik Horváth	3	áno	3
Dokumentácia k inžinierskemu dielu	Jakub Müller	5	áno	5
Dokumentácia k inžinierskemu dielu	David Silady	5	áno	5
Dokumentácia k riadeniu projektu	Táňa Poláková	5	áno	5
Dokumentácia k riadeniu projektu	Adam Šípka	5	áno	5

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	18	13	16.66667
Dominik Horváth	18	13	16.66667
Jakub Müller	10	13	16.66667
Táňa Poláková	18	13	16.66667
Dávid Silady	13	13	16.66667
Adam Šípka	12	13	16.66667

Šprint č. 4

Trvanie: 23. 11. 2021 – 07. 12. 2021

Čo hodnotíme pozitívne?

- Výsledok šprintu - naplnili sme, čo sme chceli.
- Všetky komponenty sa spolu prepletili.
- Je vidieť, že sme plánovanie robili dobre, pretože nám v tomto šprinte stačilo všetko spojiť a vracia nám to pekné výsledky.

Čo chceme zlepšiť?

- Zle sme odhadli hodiny, možno kvôli tomu, že sme body za príbehy nedávali pomocou aplikácie.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Rozbehnúť PostgreSQL	Jakub Hlavačka	3	áno	3
Komunikácia medzi aplikačným serverom a API serverom	Dávid Silady	5	áno	5
Komunikácia medzi API serverom a Elastic Search	Jakub Müller	13	áno	13
Zobrazenie výsledkov vyhľadávania	Táňa Poláková	13	áno	13
Komunikácia medzi MongoDB a API serverom	Adam Šípka	13	áno	13
Indexovanie stiahnutých dát	Jakub Hlavačka	8	áno	8
Deployment scrapera	Dominik Horváth	13	áno	13
Technický návrh developerského prostredia	Dávid Silady	8	nie	8

Riadenie projektu

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	11	13	16.66667
Dominik Horváth	13	13	16.66667
Jakub Müller	13	13	16.66667
Táňa Poláková	13	13	16.66667
Dávid Silady	13	13	16.66667
Adam Šípka	13	13	16.66667

Šprint č. 5

Trvanie: 07. 12. 2021 – 15. 12. 2021

Čo hodnotíme pozitívne?

- Nenastali problémy, lebo sme nemali príliš veľké príbehy.

Čo chceme zlepšiť?

- Všetko dopadlo dobre, tentokrát nemáme čo zlepšovať.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
[ID] Modul scraper + testovanie	Dominik Horváth	5	áno	5
[ID] Modul API server + testovanie	Jakub Müller	5	áno	5
[ID] Modul na indexovanie + testovanie	Jakub Hlavačka	5	áno	5
[ID] Modul frontend + testovanie	Táňa Poláková	5	áno	5
[ID] Modul aplikačný server + testovanie	Dávid Silady	5	áno	5
[ID] Sekv. diagram celkového fungovania	Jakub Hlavačka	1	áno	1
[RP] Metodika github actions	Adam Šípka	1	áno	1
[RP] Metodika docker	Adam Šípka	1	áno	1
[RP] Šprint č. 5	Táňa Poláková	0.5	áno	0.5
[RP] Globálna retrospektíva	Táňa Poláková	1	áno	1
[RP] Exorty úloh	Táňa Poláková	0.5	áno	0.5
Konfigurácia NGINX	Dávid Silady	1	áno	1
Testovanie dostupnosti kontajnerov	Dominik Horváth	2	áno	2
MongoDB unit test	Adam Šípka	3	áno	3

Riadenie projektu

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	6	7	16.66667
Dominik Horváth	7	7	16.66667
Jakub Müller	5	7	16.66667
Táňa Poláková	7	7	16.66667
Dávid Silady	6	7	16.66667
Adam Šípka	5	7	16.66667

Šprint č. 6

Trvanie: 24. 02. 2022 – 10. 03. 2022

Čo hodnotíme pozitívne?

- Ochotu niektorých členov tímu pracovať nad očakávaný rámec hodín (kvôli zlému odhadu).

Čo chceme zlepšiť?

- Odhad, radšej si zoberieme menej storiek a počas šprintu si pridáme, než by sme sa mali zahltiť, ako počas tohto šprintu.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Zabezpečenie endpointov cez HTTP auth basic	Jakub Müller	3	áno	3
Pridať na stránku tímu priestor na dokumenty z LS	Táňa Poláková	3	áno	3
Implementácia routera na frontende	Dávid Silady	1	áno	1
Update indexovacieho mechanizmu	Jakub Hlavačka	8	áno	10
Frontend prihlasovanie	Táňa Poláková	5	áno	5
Refaktoring MongoDB	Dominik Horváth	2	áno	2
Named entity recognition	Jakub Müller	8	áno	8
Základná user tabuľka v user databáza	Dávid Silady	5	áno	5
Stránkovanie API	Adam Šípka	13	áno	13
Implementácia developerského prostredia	Dominik Horváth	10	áno	11
Ťahanie dát (dev. prostredie)	Jakub Hlavačka	3	áno	3
Stránkovanie článkov na frontende	Táňa Poláková	5	áno	5

Riadenie projektu

Komunikácia medzi aplikač. serv. a user databázou	Dávid Silady	2	áno	7
Aktualizácia zdrojových kódov na základe zabezpečenia endpointov - API	Jakub Müller	2	zrušený	0
Aktualizácia zdrojových kódov na základe zabezpečenia endpointov - aplikačný server	Jakub Hlavačka	2	zrušený	0
Aktualizácia zdrojových kódov na základe zabezpečenia endpointov - scraper	Dominik Horváth	2	zrušený	0

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	13	13	16.66667
Dominik Horváth	13	13	16.66667
Jakub Müller	13	13	16.66667
Táňa Poláková	13	13	16.66667
Dávid Silady	8	13	16.66667
Adam Šípka	13	13	16.66667

Šprint č. 7

Trvanie: 10. 03. 2022 – 24. 03. 2022

Čo hodnotíme pozitívne?

- Upustili sme od striktného “hlasovania” o story pointoch, tým pádom bolo plánovanie efektívnejšie a rýchlejšie. Body sme pridelovali na základe spoločnej diskusie.

Čo chceme zlepšiť?

- Niektoré storky na seba nadväzovali, to sme moc nedomysleli.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Článok IIT.SRC	Adam Šípka	5	pokračuje do ďalšieho šprintu	5
Implementácia integračného testu	Jakub Müller	8	pokračuje do ďalšieho šprintu	10
Github actions na repo team web page	Dominik Horváth	1	áno	1
Autentifikácia user	Dávid Silady	8	áno	10
Vyhľadať v scrapnutom článku všetky keywords	Jakub Hlavačka	8	áno	8
Dokumentácia web frontendu	Táňa Poláková	3	toto bude aktívna storka	3
Github repo pre menšie súčasti projektu	Jakub Müller	3	áno	3
Komunikácia medzi frontendom a web serverom - poslať autentifikačné a registračné údaje	Táňa Poláková	3	pokračuje do ďalšieho šprintu	3
Komunikácia medzi aplikačným serverom a user databázou	Dávid Silady	2	áno	2

Riadenie projektu

Request na MongoDB, ktorý vráti články na základe keywords	Adam Šípka	8	pokračuje do ďalšieho šprintu	8
Integrácia nového stroja do tímového prostredia	Dominik Horváth	13	áno	13
Implementovať spôsob pamätania si usera	Táňa Poláková	5	pokračuje do ďalšieho šprintu	6
Preklad keywordov do viacerých jazykov	Jakub Hlavačka	5	áno	5
Prerobiť MongoDB	Dominik Horváth	0.5	áno	0.5

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	13	13	16.66667
Dominik Horváth	14.5	13	16.66667
Jakub Müller	11	13	16.66667
Táňa Poláková	11	13	16.66667
Dávid Silady	13	13	16.66667
Adam Šípka	13	13	16.66667

Šprint č. 8

Trvanie: 24. 03. 2022 – 07. 04. 2022

Čo hodnotíme pozitívne?

- Veľa komunikácie.
- Veľa spoločnej implementácie na frontende.

Čo chceme zlepšiť?

- Konflikty na GitHube pre application_web, musíme vymyslieť vhodný prístup.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Zobrazenie archivovaného článku	David Silady	5	nie	5
Spustenie scraperu na veľa dátach	Jakub Müller	5	áno	5
Prerobiť dev prostredie	Dominik Horváth	8	áno	8
Sidebar	David Silady	3	áno	8
Frontend keywords	Táňa Poláková	5	nie	1
Otestovať filtrovanie	Jakub Müller	8	nie	5
Prepojenie servisov	Dominik Horváth	5	áno	5
Zobraziť iba doménu stránky vo výsledkoch vyhľadávania	Táňa Poláková	1	áno	1
Frontend registrácia	Táňa Poláková	3	áno	3
PDF report z nájdených článkov	Jakub Hlavačka	13	áno	13
Filter na frontende	Táňa Poláková	5	nie	8
Implementácia integr. testu	Jakub Müller	3	áno	3
Komunikácia medzi frontendom a web serverom - poslať aut. údaje	Táňa Poláková	1	nie	1
Request na MongoDB, kt. vráti články na základe keywords	Adam Šípka	3	áno	3

Riadenie projektu

Článok IIT.SRC	Adam Šípka	8	áno	8
Rozšírenie user db	David Silady	5	nie	0
Testovať - implementovať spôsob pamätania si usera	Táňa Poláková	3	nie	3
Refactor Flask server	Adam Šípka	1	áno	2
Prerobiť mongodb	Dominik Horváth	0.5	áno	0

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	13	13	16.66667
Dominik Horváth	13.5	13	16.66667
Jakub Müller	16	13	16.66667
Táňa Poláková	14	13	16.66667
Dávid Silady	13	13	16.66667
Adam Šípka	11	13	16.66667

Šprint č. 9

Trvanie: 07. 04. 2022 – 21. 04. 2022

Čo hodnotíme pozitívne?

- Konflikty na GitHube sa výrazne zredukovali.
- Stihli sme dokončiť dôležité súčasti projektu do odovzdania testovaciemu tímu.

Čo chceme zlepšiť?

- Momentálne nám nič nenapadá.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Zobrazenie archivovaného článku	David Silady	5	áno	8
Skúsiť viac tvarov keywordov - odsledovať počet článkov	Jakub Müller	1	áno	1
Štatistika našej domény	Dominik Horváth	5	áno	5
PDF report prepojiť s prihlasovaním	Jakub Hlavačka	5	áno	5
Udržanie loginu	David Silady	3	áno	5
Frontend keywords	Táňa Poláková	5	nie	0
Otestovať filtrovanie	Jakub Müller	8	nie	8
Dokumentácia API - swagger	Dominik Horváth	3	áno	3
Kontinuálne scrapovanie	Dominik Horváth	5	áno	5
Generovanie PDF reportu	Jakub Hlavačka	8	áno	8
Filter na frontende	Táňa Poláková	3	áno	8
Testovanie flask serveru	Jakub Müller	3	áno	4
API bude vracať vetu/y, v ktorých sa nachádza hľadané slovo	Adam Šípka	5	áno	5

Riadenie projektu

Pohrať sa s query do elasticu	Adam Šípka	3	nie	3
Rozšírenie user db	David Silady	5	nie	0
About page	Táňa Poláková	5	áno	5
Rozšírenie API	Adam Šípka	5	áno	5

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	13	13	16.66667
Dominik Horváth	13	13	16.66667
Jakub Müller	13	13	16.66667
Táňa Poláková	13	13	16.66667
Dávid Silady	13	13	16.66667
Adam Šípka	13	13	16.66667

Šprint č. 10

Trvanie: 21. 04. 2022 – 18. 05. 2022

Čo hodnotíme pozitívne?

- Vhodné stanovenie cieľa šprintu - alokovali sme si čas na písanie dokumentácie, vďaka čomu sme dokázali čeliť aj neočakávanej zmene šprintu - pribudol príbeh na revíziu IIT.SRC článku.

Čo chceme zlepšiť?

- Momentálne nám nič nenapadá. Záver semestra ukončujeme úspešne.

Názov príbehu	Riešiteľ	Odhadovaný počet bodov	Je príbeh ukončený?	Pridelený počet bodov
Otestovať filtrovanie	Jakub Müller	3	áno	2
Pohrať sa s query do elasticu	Adam Šípka	1	áno	1
Plagát na IIT.SRC	Jakub Müller	8	áno	8
TP CUP - článok	Táňa Poláková	5	áno	5
TP CUP - réžia akčného videa	Dominik Horváth	5	áno	5
TP CUP - krátky opis projektu	Táňa Poláková	1	áno	1
[ID] modul elasticsearch	Dominik Horváth	5	áno	5
[ID] modul scraper	Jakub Müller	5	áno	5
[ID] modul API server	Adam Šípka	5	áno	5
[ID] modul aplikačný server	Dávid Silady	5	áno	5
[ID] modul frontend	Táňa Poláková	5	áno	5
[RP] metodika definition of ready & done	Adam Šípka	8	áno	8
[RP] big picture	Táňa Poláková	8	áno	8
[ID] big picture	Jakub Hlavačka	5	áno	5
[ID] big picture vol. 2	Adam Šípka	5	áno	5

Riadenie projektu

[RP] export úloh	Táňa Poláková	0	áno	0
[RP] popis práce v tíme	Táňa Poláková	1	áno	1
[RP] popis práce v tíme	Jakub Hlavačka	1	áno	1
[RP] popis práce v tíme	Adam Šípka	1	áno	1
[RP] popis práce v tíme	Dávid Silady	1	áno	1
[RP] popis práce v tíme	Jakub Müller	1	áno	1
[RP] popis práce v tíme	Dominik Horváth	1	áno	1
TP CUP - strihanie akčného videa	Dávid Silady	8	áno	8
IIT.SRC revízia článku pt.1	Jakub Hlavačka	13	áno	13
IIT.SRC revízia článku pt.2	Dávid Silady	5	áno	5
Príprava zdrojových kódov na odovzdanie	Dominik Horváth	8	áno	9
Logo ams	Jakub Müller	5	áno	5
Skrytie sidebaru	Dávid Silady	1	áno	1
Statická verzia stránky	Jakub Müller	1	áno	1

Meno	Súčet počtu odhadovaných bodov	Súčet počtu pridelených bodov	Podiel práce (%)
Jakub Hlavačka	19	20	16.66667
Dominik Horváth	19	20	16.66667
Jakub Müller	23	20	16.66667
Táňa Poláková	20	20	16.66667
Dávid Silady	20	20	16.66667
Adam Šípka	20	20	16.66667

Globálna retrospektíva

Zimný semester

Za zimný semester sa nám podarilo vytvoriť funkčný prototyp so statickým sťahovaním a indexovaním. Aj keď naše vízie siahali o dosť vyššie, sme s naším výkonom spokojní. Spokojnosť potvrdil aj náš vedúci tímu.

Dôvodom nenaplnenia našich pôvodných vízií môže byť do určitej miery to, že sme sa s plánovaním a vývojom projektu takéhoto rozmeru stretli prvýkrát. Pre mnohých z nás bola toto prvá skúsenosť s prácou v tíme. Veľmi nám pomohol nástroj na manažovanie úloh. Zistili sme, že pokiaľ úloha nie je zaznamenaná v nástroji, nikto ju dobrovoľne neurobí, dokonca sa na ňu zabudne. Taktiež považujeme za dôležité si zaznamenávať odpracované hodiny, pretože nám to pomáha pri plánovaní ďalších úloh.

Trvalo nám pár týždňov, kým sme sa adaptovali na fungovanie v dvojtýždňových šprintoch. Dnes sa už hrdíme piatym šprintom a môžeme konštatovať, že každý z nich bol lepší ako ten predošlý. Pri poslednom šprinte sme boli prekvapení, že sme na ňom nenašli žiadne nedostatky. Skromne si myslíme, že je to odrazom nášho neustáleho sa zlepšovania.

Po zavedení rolí v tíme sme si mysleli, že ich absencia pri prvých týždňoch vývoja projektu je pre nás veľkou stratou. Ukázalo sa, že všetko zlé je na niečo dobré. Nechali sme nás, členov tímu, aby sa rola vyvinula sama. Každý z nás postupne začal inklinovať k tomu, čo ho baví a nestretli sme sa so situáciou, že by sme rolu určili nevhodne.

Dôležitým aspektom na našej ceste k úspešnému ukončeniu tímového projektu počas zimného semestra bola fungujúca komunikácia. Je to možno tým, že sa v rámci tímu poznáme a nemáme problém povedať či zniesť konštruktívnu kritiku. Nestretli sme sa so žiadnym konfliktom záujmov.

„Najväčšou“ krízou (nebola až taká veľká) počas zimného semestra bolo písanie dokumentácie k prvému míľniku. Táto situácia sa našťastie pri druhom míľniku nezopakovala, keďže sme si na ňu vyhradili viac času.

A čo ďalej? Rozhodne budeme pokračovať v zabehnutých koľajniciach našich šprintov. V letnom semestri očakávame zväčšenie úložiska pre náš projekt, čo výrazne ovplyvní jeho funkčnosť. Domnievame sa, že tú najťažšiu robotu už máme za sebou a odteraz to už bude len nadstavba funkcionality, na ktorú sa veľmi tešíme. Náš projekt má zmysel, a to je našou každodennou motiváciou.

Letný semester

V rámci piatich šprintov sme dali do pohybu kontinuálne scrapovanie článkov, ktorých počet presahuje 3 milióny. Prototyp sme úspešne a včas odovzdali na testovanie. Implementovali nové funkcionality, avšak na pár z nich neostal čas.

Prvý týždeň letného semestra manažér plánovania kriticky zhodnotil, čo by sme v priebehu šprintov, ktoré nás čakali, mohli stihnúť. Tieto ciele sme prioritizovali na základe konzultácie

Riadenie projektu

s vedúcim tím. Myslíme si, že práve tento krok viedol k úspešnému odovzdaniu prototypu na testovanie. Plánovanie tohto rozmeru v zimnom semestri nebolo možné v zimnom semestri, keďže logika fungovania celého projektu sa v tomto období ešte len kryštalizovala.

Ako v dokumente uvádzame, často sme čelili nesprávnemu odhadu náročnosti príbehov. Tento problém sa nám čiastočne podarilo odstrániť vďaka upusteniu od hlasovania o body, ktoré sme príbehom udeľovali. Spočiatku by sa mohlo zdať, že viac ľudí odhadne počet bodov lepšie, ale v našom prípade bol opak pravdou. O počte bodov vždy rozhodol človek, ktorému bol daný príbeh pridelený. Príbehy sa zväčša prideliili na základe skúseností členov tímu.

Projekt sme predstavili aj na IIT.SRC, ale z hľadiska riadenia projektu sme náročnosť písania článku podcenili. Je to už druhýkrát (prvýkrát v zimnom semestri - kríza pri prvom míľniku), kedy nám problém spôsobila úloha spojená s písaním. Uvedomujeme si, že v úlohách takéhoto charakteru nás zachráni iba priebežná práca.

Našou víziou bolo aj aplikovanie pokročilejších informatických technológií do projektu. Dôvodom, prečo sme sa k nim nedostali, môže byť sekvenčný prístup k implementácii jednotlivých funkcionalít. Tým, že náš tím je z technologického hľadiska rôznorodý, by možno stálo za úvahu, pristúpiť k implementácii funkcionalít paralelne. Potenciálne by to mohlo celý vývoj urýchliť. Mali by sme na výber z viacerých oblastí, v ktorých si môžeme vymyslieť úlohu/príbeh a minimalizovali by sa prípady, kedy by člen tímu dostal iba nejaký zvyškový príbeh. Takéto situácie vyvolávali značnú demotiváciu a práca na príbehoch trvala spravidla dlhšie.

Napriek všetkým negatívam sme s našou prácou počas celého akademického roka spokojní. Pripravili sme základy pre projekt, ktorý má potenciál byť vyvíjaný aj v budúcnosti a nadobudli sme množstvo technologických aj manažérskych vedomostí a schopností.

Motivačný dokument

Tím

Náš tím tvoria perspektívni ľudia, ktorých záujmy pokrývajú rôzne oblasti informatiky. Poznáme sa už od strednej školy. Počas štúdia sme spolupracovali na rôznych projektoch, čo nás naučilo hľadať efektívne kompromisy. Zameriavame sa najmä na dátovú vedu, webové technológie a umelú inteligenciu, čomu zodpovedajú aj naše absolvované predmety na bakalárskom stupni štúdia (Vývoj aplikácií v jazyku JavaScript, Webové technológie, Inteligentná analýza údajov). Niektorí z nás si tieto vedomosti plánujú rozšíriť na predmetoch Vývoj webových aplikácií v prostredí cloud a Neurónové siete.

Väčšina členov nášho tímu aktuálne pôsobí v zamestnaní, ktoré im dáva cennú informatickú prax. Dávid Silady a Dominik Horváth sú kolegovia vo firme IBM, kde pracujú na pozíciach Package specialist a DevOps inžinier. Automatizácia procesov, Docker, Kubernetes, Bash alebo Python sú častou súčasťou ich práce. Táňa Poláková má hlavnú náplň práce zber a predspracovanie veľkého množstva dát (využíva MySQL, numpy), vďaka ktorým bude môcť navrhnúť model neurónovej siete, ktorý nahradí aktuálne riešenie. Jakub Hlavačka v práci často využíva technológie Scrapy a Selenium na scrapovanie údajov z internetu.

Naše bakalárske práce sa zaoberali dátovou analýzou, umelou inteligenciou, technológiou blockchain a webom. Veľkou hodnotou (tak, ako aj ostatní členovia) v našom tíme je Dominik Horváth, ktorého bakalársky projekt bol zameraný na dátovú analýzu falošných správ súvisiacich s pandemiou COVID-19, a jeho práca bola jednou z prvých, ktoré sa venovali tejto oblasti. Daná tematika sa priamo dotýka témy na tímový projekt, o ktorý máme najväčší záujem. Táňa Poláková a Dávid Silady sa na svojich bakalárkach zaoberali neurónovými sieťami (pracovali s knižnicami TensorFlow a PyTorch). Dávid odhadoval hĺbku z fotografií pomocou neurónových sietí, kde sa zoznámil najmä s generovaním a spracovaním dát, rovnako ako so základmi konvolučných sietí. Táňa trénovala neurónovú sieť, ktorej výstup nezávisel od usporiadania prvkov vo vstupe, pričom vstup boli medicínske dáta. Práca Jakuba Hlavačku bola zameraná na využitie technológie blockchain v IoT zariadeniach, kde nadobudol prvé väčšie znalosti so systémami ako Docker a Hyperledger Fabric. Jakub Müller má zo svojej bakalárskej práce skúsenosti z oblasti tvorby webových stránok a UX testovania, keďže jeho práca bola zameraná na používateľmi vnímanú dôveryhodnosť v online spravodajstve. Adam Šípka v bakalárke vytvoril webový crawler (použil framework Selenium), ktorý zaznamenával a ukladal metadáta o YouTube videách. Vďaka tomu získal skúsenosti so získavaním dát a následne aj s ich prespracovaním a analýzou.

Každý z nás má vášeň pre web. Dávid obľubuje prácu s webovými technológiami ako React, Nextjs a Express(Nodejs). Dominik preferuje frameworky ako Flask alebo Django a baví ho bezpečnosť webových technológií. Táňa sa rada kreatívne vyžije pri tvorbe webového front-endu (pomocou frameworku React), avšak má skúsenosti aj s back-endom (Nodejs). Jakub H. sa najviac zaujíma o vytváranie webových aplikácií s prístupom Fullstack za pomoci technológií ako Express, React, Flask, Bootstrap, SQL. Jakub M. sa najradšej angažuje v oblasti front-endu a má skúsenosti s tvorbou webových aplikácií. Adam sa v jednom zo svojich projektov venoval parsovaniu dát z rôznych webových stránok.

Motivácia

Téma č. 4 - Adverse Media Screening [AMS]

Tému Adverse Media Screening [AMS] sme si vybrali najmä kvôli zaujímavosti a rôznorodosti infromatických úloh, ktoré budú pri jej vývoji potrebné. Vďaka väčšinovému zameraniu tímu na webové technológie nám nebude robiť problém tvorba front-end a back-end rozhrania. Niektorí členovia tímu majú taktiež skúsenosti so “scrapingom” rôznych webových portálov (pomocou technológií Scrapy a Selenium) na ďalšie spracovanie dát z nich, čo je pri tejto téme dôležitá skúsenosť.

Každý člen tímu má v prvom semestri inžinierskeho ročníka predmety Vyhľadávanie informácií a Neurónové siete, čo nám výrazne pomôže pri problematike vybranej témy. Svoju kvalifikáciu doplníme absolvovanými predmetmi ako Inteligentná analýza údajov, Webové technológie a Vývoj aplikácií v JavaScript.

Zameranie bakalárskej práce dvoch členov nášho tímu bolo na neurónové siete a jeden z členov pracoval na podobnej téme ako je téma AMS vo svojej bakalárskej práci, kde využil technológiu Elasticsearch pre vyhľadávanie informácií vo veľkej dátovej sade.

Téma č. 8 - Educational Content Engineering Hub - Databáza otázok, odpovedí, úloh a riešení [ECEH-DU]

Táto téma sa taktiež z veľkej časti opiera o technológie, s ktorými sme všetci nielen stotožnení, ale naozaj nás bavia. Tiež sa jedná o oblasť, ktorá nám je ako študentom veľmi blízka, a bolo by zaujímavé prispieť do procesu testovania svojou prácou, ktorá potenciálne môže priniesť niečo nové. Skúsenosti tímu s webovými technológiami, DevOps a taktiež poznatky z prostredia veľkých projektov v pracovnom svete budú v tomto projekte veľmi cenné.

Príloha A - Zoradenie všetkých tém podľa priority

V tabuľke nižšie sa nachádza zoradenie 12-ich tém podľa priority.

Priorita témy	Číslo témy	Názov témy
1.	4.	Adverse Media Screening [AMS]
2.	8.	Educational Content Engineering Hub - Databáza otázok, odpovedí, úloh a riešení [ECEH-DU]
3.	16.	FIIT WIX
4.	17.	Document Wizard
5.	3.	DataHub pre rôzne typy zariadení, ich spracovanie / analýzu / vizualizáciu
6.	11.	(Q)SAR analýza fototoxických látok
7.	15.	Ion Mobility Spectrometry for Rapid HEMP Potency Testing
8.	14.	IoT platforma na priemyselnú automatizáciu - malý pivovar
9.	2.	Transformácia priestorov na bezpečné a inteligentné miesta na prácu [space2]
10.	9.	Monitorovanie a správa systému pre výrobný areál [LOMON]
11.	18.	Webové IDE pre ASIC [ASICDE]
12.	19.	Automatizácia procesov KYC (Know your client) a AML [Anti-money laundering]

Príloha B - Rozvrh tímu

V prílohe sa nachádzajú povinnosti jednotlivých členov nášho tímu. Stretnutia tímu s vedúcim a spoločnú prácu tímu sme stanovili nasledovne:

- Stretnutie tímu s vedúcim: Utorok 11:00 - 14:00
- Spoločná práca tímu: Pondelok 15:00 - 18:00

Sme otvorení iným možnostiam na stretnutie s vedúcim v prípade, že by náš potenciálny vedúci nebol v náš preferovaný čas k dispozícii.

Dominik Horváth

Deň	8.00-8.50	9.00-9.50	10.00-10.50	11.00-11.50	12.00-12.50	13.00-13.50	14.00-14.50	15.00-15.50	16.00-16.50	17.00-17.50	18.00-18.50	19.00-19.50
Po	1.28 (LIS) (BA-FIIT-FIIT) Vyhľadavanie informácií (1) M. Kováč								Spoločná práca tímu			1.38 (U20b) (BA-FIIT-FIIT) Architektúra softvéru (1) L. Graf
Ut	Stretnutie tímu s vedúcim						-1.61 (Aula Magna) (BA-FIIT-FIIT) Architektúra softvéru V. Vranic	-1.61 (Aula Magna) (BA-FIIT-FIIT) Výskum inteligentných softvérových systémov V. Vranic			-1.61 (Aula Magna) (BA-FIIT-FIIT) Timový projekt I (2) K. Košťál	
St			-1.58 (U120) (BA-FIIT-FIIT) Vyhľadavanie informácií (3) M. Kováč					-1.61 (Aula Magna) (BA-FIIT-FIIT) Manažment v tvorbe softvéru (3) I. Černáková				1.29 (LWT) (BA-FIIT-FIIT) Manažment v tvorbe softvéru (3) I. Černáková
St	-1.58 (U120) (BA-FIIT-FIIT) Základy kryptografie V. Khylenko		1.30a (LSS1) (BA-FIIT-FIIT) Základy kryptografie V. Khylenko									
Pi												

Jakub Hlavačka

Deň	8.00-8.50	9.00-9.50	10.00-10.50	11.00-11.50	12.00-12.50	13.00-13.50	14.00-14.50	15.00-15.50	16.00-16.50	17.00-17.50	18.00-18.50	19.00-19.50	
Po			1.28 (LIS) (BA-FIIT-FIIT) Vyhľadavanie informácií (1) M. Kováč								Spoločná práca tímu		1.38 (U20b) (BA-FIIT-FIIT) Architektúra softvéru (1) L. Graf
Ut	Stretnutie tímu s vedúcim						-1.61 (Aula Magna) (BA-FIIT-FIIT) Architektúra softvéru V. Vranic	-1.61 (Aula Magna) (BA-FIIT-FIIT) Výskum inteligentných softvérových systémov V. Vranic			-1.61 (Aula Magna) (BA-FIIT-FIIT) Timový projekt I (2) K. Košťál		
St			-1.58 (U120) (BA-FIIT-FIIT) Vyhľadavanie informácií (3) M. Kováč					-1.61 (Aula Magna) (BA-FIIT-FIIT) Manažment v tvorbe softvéru (3) I. Černáková				1.28 (LIS) (BA-FIIT-FIIT) Manažment v tvorbe softvéru (3) F. Lehocki	
St	-1.58 (U120) (BA-FIIT-FIIT) Základy kryptografie V. Khylenko		1.30a (LSS1) (BA-FIIT-FIIT) Základy kryptografie V. Khylenko										
Pi													

Táňa Poláková

Deň	8.00-8.50	9.00-9.50	10.00-10.50	11.00-11.50	12.00-12.50	13.00-13.50	14.00-14.50	15.00-15.50	16.00-16.50	17.00-17.50	18.00-18.50	19.00-19.50	
Po									Spoločná práca tímu			1.38 (U20b) (BA-FIIT-FIIT) Architektúra softvéru (1) L. Graf	
Ut	1.28 (LIS) (BA-FIIT-FIIT) Vyhľadavanie informácií M. Kováč		Stretnutie tímu s vedúcim						-1.61 (Aula Magna) (BA-FIIT-FIIT) Architektúra softvéru V. Vranic	-1.61 (Aula Magna) (BA-FIIT-FIIT) Výskum inteligentných softvérových systémov V. Vranic			-1.61 (Aula Magna) (BA-FIIT-FIIT) Timový projekt I (2) K. Košťál
St			-1.58 (U120) (BA-FIIT-FIIT) Vyhľadavanie informácií (3) M. Kováč					-1.61 (Aula Magna) (BA-FIIT-FIIT) Manažment v tvorbe softvéru (3) I. Černáková				1.29 (LWT) (BA-FIIT-FIIT) Manažment v tvorbe softvéru (3) I. Černáková	
St	-1.58 (U120) (BA-FIIT-FIIT) Základy kryptografie V. Khylenko		1.30a (LSS1) (BA-FIIT-FIIT) Základy kryptografie V. Khylenko										
Pi													

Jakub Müller

Metodiky

Metodika Scrum

Na riadenie projektu využívame metodiku SCRUM, ktorú sme si prispôbili pre potreby nášho projektu. Dodržiavanie princípov má na starosti SCRUM master, ktorý okrem toho prideluje úlohy ostatným členom tímu v nástroji Jira a celkovo spravuje tento nástroj. Naše šprinty trvajú dva týždne a začínajú v utorok, v deň stretnutia s vedúcim tímu.

Standup

Každý pondelok, deň pred stretnutím s vedúcim, prebehne stretnutie tímu, kde sa všetci členovia tímu vyjadria ako prebieha plnenie im pridelených úloh, s primárnym zameraním na problémy, ktoré sa vyskytli. Následne prebehne diskusia s cieľom vyriešiť identifikované problémy.

Šprint review

V posledný deň šprintu, pri stretnutí s vedúcim (v utorok), sa vykoná šprint review, kde sa na začiatku stretnutia každý člen vyjadrí k svojim úlohám, ako sa mu ich podarilo splniť, prípadne opísať problémy, ktoré sa pri plnení úlohy vyskytli. Cieľom je vyvolať diskusiu, ktorá by nám pomohla posunúť sa ďalej a vyriešiť daný problém. Ak úloha ešte nebola vyriešená, tak sa prenesie do nasledujúceho šprintu. Okrem toho každý člen povie, koľko času na daných úlohách strávil. Strávený čas sa následne porovná s odhadovaným časom, čo nás naučí tvoriť presnejšie odhady.

Po vyjadreniach od všetkých členov sa zhodnotí samotný šprint ako celok. Členovia sa vyjadria, čo hodnotia ako pozitívne, alebo naopak, v akých oblastiach sa potrebujú zlepšiť.

Plánovanie

V deň stretnutia, po tom, čo prebehne šprint review, sa naplánujú úlohy na nasledujúci šprint. Novým príbehom prideliujeme kartičky s hodnotami 0.5, 1, 2, 3, 5, 8, 13 a ?, ktoré reprezentujú náročnosť úlohy. V našom prípade sa náročnosť rovná počtu predpokladaných hodín, ktoré daný člen pri riešení úlohy odpracuje. Takto postupujeme kvôli ľahšiemu plánovaniu. Cieľom je, aby každý člen odpracoval za jeden šprint 12 hodín (6 na týždeň). Táto hodnota nemusí byť nutne dodržaná, je možné, že v niektorých prípadoch bude úloha vyriešená za kratší, alebo dlhší čas.

Pri prideľovaní sa uprednostňujú úlohy, ktoré nás posunú bližšie k cieľu projektu. Ak nastane situácia, kde momentálne nebudeme schopný vytvoriť úlohu a niektorý z členov bude mať pridelenú prácu s príliš nízkym počtom hodín, bude mu priradená úloha z backlogu s najvyššou prioritou.

Práca v nástroji Jira

Pri práci používame nástroj Jira, s ktorým pracuje primárne náš SCRUM master. Pridelované a evidované tasky sú rozdelené do nasledovných kategórií:

Riadenie projektu

- Biznis analýza
- Front end
- Back end - server
- Back end - data
- Softvérová architektúra
- Stránka tímu

Práca v nástroji Jira je podrobnejšie popísaná v nasledujúcich sekciách.

Backlog

V backlogu sa nachádzajú tasky, o ktorých vieme, že v budúcnosti sa budú musieť implementovať, no v momentálnej dobe nie sú prioritné. Pridávanie nových taskov do backlogu prebieha na stretnutiach tímu, alebo na stretnutiach s vedúcim. Ak vieme dopredu identifikovať požadovanú funkcionálnu našo riešenia a čo bude pre ňu potrebné analyzovať/implementovať, tak pre tieto problémy vytvoríme tasky, ktoré budú umiestnené v backlogu. Počas plánovania nového šprintu môže byť niektorá z už pridelených úloh taktiež presunutá do backlogu, ak sa všetci členovia zhodnú, že momentálne existujú problémy s vyššou prioritou.

Nástenka

Na nástenke sú zobrazené tasky aktuálneho šprintu. V prvom semestri boli rozdelené do troch stĺpcov podľa ich aktuálneho stavu riešenia (*to do*, *in progress* a *done*). Nasledujúci semester sme nástenku rozšírili, boli pridané dva nové stĺpce: *to test* a *to review*. Na ich lepšie pochopenie slúžia dokumenty *Definition of ready* a *Definition of done*.

Pri každom tasku je v detailoch uvedený odhadovaný čas. Člen tímu, ktorému je daná úloha pridelená priebežne aktualizuje odrobený čas, ktorý sa pri šprint review bude porovnávať s odhadom. Okrem toho do komentára k danému tasku napíše krátke zhrnutie jeho výsledkov, prípadne problémov, ktoré sa vyskytli.

Git Metodika

Na verziovanie zdrojového kódu je používaný Git. Náš tím má na GitHube vytvorenú organizáciu FIIT-TEAM8, kde sú v jednotlivých repozitároch uložené zdrojové kódy danej časti projektu. Aktuálny zoznam repozitárov je nasledovný:

- application web
- script na prepojenie Elasticsearch s Mongo databázou
- webová stránka tímu
- scraper napojený na Mongo databázu
- flask server

Pri verziovaní kódu sú uplatnené praktiky, ktoré opisujeme v nasledujúcich sekciách.

Vetvenie

Main branch je určená na aktuálnu verziu funkčného kódu. Nie je určená na samotný vývoj alebo implementáciu nových funkcií, tie prebiehajú v oddelených vetvách, ktoré sú vytvorené členom tímu, ktorý na dane úlohe pracuje. Meno takejto vetvy musí popisovať cieľ implementovanej funkcionality. Vďaka vetveniu môžu na rovnakej časti projektu pracovať viacerí členovia tímu súčasne a bez toho, aby došlo k veľkým konfliktom v zdrojovom kóde.

Commit

Commity si vytvára každý člen tímu na svojej vetve, ale napriek tomu, by mali mať zmysluplný dôvod a mali by byť tvorené až po implementovaní určitej časti kódu, alebo opravení zistených chýb. Commit message je písaný v anglickom jazyku a je dostatočne deskriptívny, no bez toho, aby bol príliš rozsiahly.

Pull requests

Po úspešnom implementovaní funkcionality, alebo odstránení problému môže člen tímu vytvoriť pull request, aby aktualizoval kód v repozitári. Merge konflikty, ktoré sa v takýchto situáciách môžu niekedy vyskytnúť, je nutné pred uzavretím pull requestu odstrániť bez toho, aby utrpela funkcionality programu.

Metodika nasadzovania

Docker

Väčšina modulov, opísaných v sekcii Moduly systému, je nasadených prostredníctvom technológie Docker, čo nám uľahčuje stavanie, nasadzovanie a spravovanie jednotlivých modulov. Každý z modulov musí byť umiestnený vo vlastnom samostatnom kontajneri, keďže sa neodporúča, aby v jednom kontajneri boli obsiahnuté moduly zaoberajúce sa rozdielnymi oblasťami. Toto nám umožňuje zabaliť našu aplikáciu so všetkými jej súčasťami, ako sú knižnice, databázy, dependencies, a následne ju nasadiť na viacerých platformách.

Momentálne sú využívané nižšie uvedené kontajnery:

- nginx: webový server
- es01: kontajner pre ElasticSearch
- mongo_db: databáza článkov
- postgres_db: user databáza
- mongo_express
- flask_server: kontajner pre API server
- node_server: kontajner pre webový server
- scraper
- index_service

Komunikácia kontajnerov

Niektoré z týchto kontajnerov dokážu medzi sebou komunikovať prostredníctvom virtuálnej siete pomenovanej *elastic_mongo*. Konkrétne sa jedná o kontajnery *index_service*, *mongo_db* a *es01*. Už z názvu sa dá poznať, že sa jedná o prepojenie ElasticSearch-u a Mongo databázy. Vďaka tejto sieti sa môžu indexované dáta hneď ukladať do databázy.

Vytvárať takéto siete je nutné, keďže jednotlivé komponenty sú izolované v samostatných kontajneroch, no stále budú musieť medzi sebou vedieť komunikovať, čiže navzájom si posilať, alebo prijímať requesty.

GitHub Actions

Pri projekte používame GitHub Actions, čo je platforma nepretržitej integrácie a nepretržitého doručovania (ang. continuous integration and continuous delivery, CI/CD), ktorá umožňuje automatizovať zostavenie, testovanie a následné nasadzovanie nášho riešenia. Vďaka nej môžeme vytvárať tzv. workflows, čo sú pracovné postupy, ktoré zostavujú a testujú každý pull request do repozitára, alebo nasadzujú merged pull request do produkcie.

Build & Deploy

GitHub Actions sú konkrétne využité na automatické zostavenie projektu, ktoré prebehne, vždy, keď niektorý z členov tímu vytvorí pull request v niektorom z repozitárov našej GitHub organizácie. Momentálne je táto akcia implementovaná len v repozitároch *flask_server*, *elasticsearch_mongo* a *application_web*. V tomto kroku je vytvorený docker image, ktorý je následne nahratý na Docker Hub.

Riadenie projektu

Okrem toho GitHub Actions používame na nasadzovanie zostavených projektov do produkcie. Tento krok prebehne vždy, keď bude pull request zlúčený s hlavnou vetvou programu. Aby došlo k nasadeniu, je najprv nutné pripojiť sa k našej virtuálnej mašine, kde beží Docker engine. Toto sa vykonáva prostredníctvom SSH. Na záver sa v tejto lokácii spustí Docker kontajner. Vďaka tejto funkcionalite nie je nutné náš projekt zostavovať a nasadzovať manuálne, čo nám výrazne šetrí čas.

Definition of ready

Predtým ako môže byť task považovaný za ready, je nutné splniť niekoľko dole uvedených podmienok. Až ak sú tieto podmienky splnené, task sa môže presunúť z panelu *to test* do panelu *to review*. Konkrétne sa jedná o:

- funkčný a otestovaný kód
- bol otvorený pull request
- k danému tasku je pridelený code reviewer

Následne pridelený code reviewer vykoná code review. Okrem hľadania chýb code reviewer posudzuje aj efektívnosť danej implementácie a prehľadnosť kódu. Kedykoľvek môže kontaktovať autora kódu, aby mu vysvetlil prípadné nejasnosti. Identifikované problémy môže riešiť dvoma hlavnými spôsobmi. Ak sa jedná o chybu malého rozsahu, opravu vykoná on sám. Ak je nájdený problém mohutnejší, code reviewer znova kontaktuje autora a objasní mu danú situáciu. V takýchto prípadoch je task presunutý naspäť do panelu *in progress*.

Definition of done

Task sa považuje za dokončený, až keď sú splnené dole uvedené podmienky. Po ich splnení môže byť presunutý z panelu *to review* do panelu *done*. Jedná sa o:

- úspešný code review
- uzatvorený pull request
- úspešné nasadenie do produkcie
- aktualizovaná dokumentácia

Ak by náhodou nastala situácia, kde po nasadení do produkcie (uzatvorený pull request) nebude riešenie fungovať, daný task sa vráti do panelu *in progress*.

Export evidencie úloh

V nasledujúcej tabuľke sa nachádza evidencia jednotlivých príbehov k dňu ukončenia posledného šprintu (18. 05. 2022). Export obsahuje všetky doposiaľ plánované, ukončené, zrušené a aktuálne plnené príbehy. Body sa nenachádzajú pri každom príbehu, nakoľko niektoré z nich sú v backlogu a niektoré z nich vznikli ešte v čase, kedy sme body nepoužívali.

Issue Type	Summary	Story point estimate	Assignee	Status	Created	Resolved	Sprint	Sprint_1
Story	[ID] big picture vol:2	5	Adam Šipka	Done	05-05-22 11:38	18-05-22 11:24	AMS Sprint 10	
Story	Statická verzia stránky	1	Jakub Müller	Done	05-05-22 11:35	17-05-22 23:01	AMS Sprint 10	
Story	Upraví mierne posunutie stránky na mobile			To Do	05-05-22 11:30			
Story	Skrytie sidebaru	1	David Slady	Done	05-05-22 11:29	18-05-22 11:24	AMS Sprint 10	
Story	Logo ams	5	Jakub Müller	Done	05-05-22 11:27	18-05-22 11:24	AMS Sprint 10	
Story	Príprava zdrojových kódov na odovzdanie	8	Dominik Horvath	Done	05-05-22 11:20	18-05-22 11:24	AMS Sprint 10	
Story	TP CUP - strihanie akčného videa	8	David Slady	Done	02-05-22 9:07	08-05-22 16:14	AMS Sprint 10	
Story	IIT_SRC revízia článku pt.2	5	David Slady	Done	02-05-22 9:03	02-05-22 9:03	AMS Sprint 10	
Story	IIT_SRC revízia článku pt.1	13	Jakub Hlavačka	Done	02-05-22 9:03	02-05-22 9:03	AMS Sprint 10	
Story	Zobrazenie regiónu, odkiaľ článok pochádza			To Do	21-04-22 12:06			
Story	Stress test			To Do	21-04-22 12:01			
Story	[RP] export úloh	0	Táňa Poláková	Done	21-04-22 11:48	18-05-22 11:24	AMS Sprint 10	
Story	[RP] big picture	8	Táňa Poláková	Done	21-04-22 11:48	08-05-22 16:14	AMS Sprint 10	
Story	[RP] Popis prác v time - Dávid Slady	1	David Slady	Done	21-04-22 11:47	18-05-22 10:25	AMS Sprint 10	
Story	[RP] Popis prác v time - Jakub Müller	1	Jakub Müller	Done	21-04-22 11:46	18-05-22 10:39	AMS Sprint 10	
Story	[RP] Popis prác v time - Adam Šipka	1	Adam Šipka	Done	21-04-22 11:46	29-04-22 9:23	AMS Sprint 10	
Story	[RP] Popis prác v time - Jakub Hlavačka	1	Jakub Hlavačka	Done	21-04-22 11:46	18-05-22 10:38	AMS Sprint 10	
Story	[RP] Popis prác v time - Táňa Poláková	1	Táňa Poláková	Done	21-04-22 11:46	28-04-22 21:39	AMS Sprint 10	
Story	[RP] Popis prác v time - Dominik Horvath	1	Dominik Horvath	Done	21-04-22 11:45	18-05-22 10:39	AMS Sprint 10	
Story	[ID] big picture	5	Jakub Hlavačka	Done	21-04-22 11:43	18-05-22 11:24	AMS Sprint 10	
Story	[RP] metodika definition of ready & done	8	Adam Šipka	Done	21-04-22 11:42	29-04-22 14:17	AMS Sprint 10	
Story	[ID] modul frontend	5	Táňa Poláková	Done	21-04-22 11:41	18-05-22 11:20	AMS Sprint 10	
Story	[ID] modul aplikčný server	5	David Slady	Done	21-04-22 11:41	18-05-22 10:53	AMS Sprint 10	
Story	[ID] modul API server	5	Adam Šipka	Done	21-04-22 11:40	29-04-22 16:49	AMS Sprint 10	
Story	[ID] modul scraper	5	Jakub Müller	Done	21-04-22 11:40	18-05-22 11:24	AMS Sprint 10	
Story	[ID] modul elasticsearch	5	Dominik Horvath	Done	21-04-22 11:39	18-05-22 11:24	AMS Sprint 10	
Story	TP CUP - krátky opis projektu	1	Táňa Poláková	Done	21-04-22 11:35	08-05-22 16:14	AMS Sprint 10	
Story	TP CUP - režia akčného videa	5	Dominik Horvath	Done	21-04-22 11:34	08-05-22 16:14	AMS Sprint 10	
Story	TP CUP - Článok	5	Táňa Poláková	Done	21-04-22 11:34	08-05-22 16:14	AMS Sprint 10	
Story	Plagát na IIT_SRC	8	Jakub Müller	Done	21-04-22 11:33	02-05-22 9:05	AMS Sprint 10	
Story	Štatistika našej domény	5	Dominik Horvath	Done	07-04-22 12:12	16-04-22 12:09	AMS Sprint 9	
Story	API bude vracat vetu/y, v ktorých sa nachádza hľadane slovo	5	Adam Šipka	Done	07-04-22 12:06	21-04-22 11:29	AMS Sprint 9	

Story	Testovanie flask serveru	5	Jakub Müller	Done	07-04-22 11:59	21-04-22 11:28	AMS Sprint 9	
Story	Pohraj sa s query do elasticu	3	Adam Šipka	Done	07-04-22 11:57	02-05-22 19:19	AMS Sprint 10	
Story	Rozšírenie API	5	Adam Šipka	Done	07-04-22 11:51	21-04-22 11:28	AMS Sprint 9	
Story	About page	5	Táňa Poláková	Done	07-04-22 11:39	21-04-22 11:28	AMS Sprint 9	
Story	Dokumentácia API - swagger	3	Dominik Horvath	Done	07-04-22 11:37	16-04-22 12:05	AMS Sprint 9	
Story	Udržanie logínu	3	David Slady	Done	07-04-22 11:32	17-04-22 21:11	AMS Sprint 9	
Story	Skúsiť viac tvarov keywordov - odsledovať počet článkov	1	Jakub Müller	Done	07-04-22 11:29	21-04-22 0:30	AMS Sprint 9	
Story	Kontinuálne scrapovanie	5	Dominik Horvath	Done	07-04-22 11:27	16-04-22 12:09	AMS Sprint 9	
Story	Generovanie PDF reportu	8	Jakub Hlavačka	Done	07-04-22 11:26	21-04-22 11:28	AMS Sprint 9	
Story	PDF report prepojiť s priradením	5	Jakub Hlavačka	Done	07-04-22 11:24	21-04-22 11:27	AMS Sprint 9	
Story	Statistika domén, ktoré vracia Google News - kolko ich je, početnosť vrátania článkov			To Do	31-03-22 11:06			
Story	Refactor Flask server		Adam Šipka	Done	25-03-22 18:00	07-04-22 11:31	AMS Sprint 8	
Story	Náhľad článku			To Do	24-03-22 12:24			
Story	Zobrazenie archivovaného článku	5	David Slady	Done	24-03-22 12:21	21-04-22 11:23	AMS Sprint 9	
Story	Spustenie scraperu na veľa dátach	5	Jakub Müller	Done	24-03-22 12:20	07-04-22 11:27	AMS Sprint 8	
Story	Prerobiť developerské prostredie	8	Dominik Horvath	Done	24-03-22 12:12	07-04-22 11:27	AMS Sprint 8	
Story	Sidebar	3	David Slady	Done	24-03-22 12:09	07-04-22 11:31	AMS Sprint 8	
Story	Frontend keywords prekladanie pre používateľa	5	Táňa Poláková	To Do	24-03-22 11:43			
Story	Otestovať filtrovanie	8	Jakub Müller	Done	24-03-22 11:20	04-05-22 22:29	AMS Sprint 10	
Story	Prepojenie servisov	5	Dominik Horvath	Done	24-03-22 10:57	07-04-22 11:27	AMS Sprint 8	
Story	Profil používateľa			To Do	24-03-22 10:52			
Story	Uppdate user account data		David Slady	To Do	24-03-22 10:50			
Story	Zobraziť iba doménu stránky vo výsledkoch vyhľadávania	1	Táňa Poláková	Done	20-03-22 9:31	07-04-22 11:04	AMS Sprint 8	
Story	Prerobiť mongodb	0.5	Dominik Horvath	Done	17-03-22 11:49	19-03-22 20:59	AMS Sprint 7	
Story	Implementácia integráčného testu	8	Jakub Müller	Done	10-03-22 11:50	07-04-22 11:23	AMS Sprint 8	
Story	Komunikácia medzi frontendom a web serverom - poslať autentifikačné a registračné údaje	3	Táňa Poláková	Done	10-03-22 11:44	14-04-22 21:50	AMS Sprint 9	
Story	Implementovať spôsob pamätania si usera	5	Táňa Poláková	Done	10-03-22 11:41	07-04-22 11:28	AMS Sprint 8	
Story	Menu pre každý zobrazený článok			To Do	10-03-22 11:30			
Story	Dokumentácia web frontendu	3	Táňa Poláková	Done	10-03-22 11:26			
Story	Integrácia nového stroja do tímového prostredia	13	Dominik Horvath	Done	10-03-22 11:25	24-03-22 11:21	AMS Sprint 7	
Story	Github actions na repo team web page	1	Dominik Horvath	Done	10-03-22 11:25	19-03-22 21:23	AMS Sprint 7	
Story	Github repo pre mensie susasti projektu	3	Jakub Müller	Done	06-03-22 22:12	20-03-22 13:52	AMS Sprint 7	
Story	Request na MongoDB, ktorý vráti články na základe keywords	8	Adam Šipka	Done	03-03-22 11:54	07-04-22 11:31	AMS Sprint 8	

Story	Boolean query implementácia				To Do	03-03-22 11:48			
Story	Vyhľadať v scrapuatom článku všetky keywords	8	Jakub Hlavačka	Done	03-03-22 11:40	24-03-22 11:21	AMS Sprint 7		
Subtask	káhanie dát	3	Jakub Hlavačka	Done	24-02-22 12:38	03-03-22 11:30	AMS Sprint 6		
Story	Implementácia developerského prostredia	13	Dominik Horvath	Done	24-02-22 12:38	03-03-22 11:30	AMS Sprint 6		
Story	Implementácia routera na frontende	1	David Slady	Done	24-02-22 12:21	11-03-22 16:06	AMS Sprint 6		
Story	Google analytics			To Do	24-02-22 12:18				
Story	Obmedzenie API pre tretiu stranu			To Do	24-02-22 12:02				
Story	Previdencia zahŕtenia API			To Do	24-02-22 11:26				
Story	PDF report zdieľateľný linkom			To Do	24-02-22 11:08				
Story	Dokumentácia - metodika definition of done			To Do	19-02-22 11:04				
Story	Dokumentácia - metodika definition of ready			To Do	19-02-22 11:04				
Story	Pridať na stránku tímu priestor na dokumenty z LS	3	Táňa Poláková	Done	19-02-22 10:57	03-03-22 11:06	AMS Sprint 6		
Story	Preklad keywordov do viacerých jazykov	5	Jakub Hlavačka	Done	19-02-22 10:51	24-03-22 11:18	AMS Sprint 7		
Story	Článok IIT:SRC	5	Adam Šipka	Done	19-02-22 10:46	07-04-22 11:27	AMS Sprint 8		
Story	Named entity recognition	8	Jakub Müller	Done	19-02-22 10:46	11-03-22 16:08	AMS Sprint 6		
Story	Frontend registrácia	3	Táňa Poláková	Done	19-02-22 10:44	07-04-22 11:25	AMS Sprint 8		
Story	Frontend obrazovka prihlásenia	5	Táňa Poláková	Done	19-02-22 10:40	11-03-22 16:08	AMS Sprint 6		
Story	Update indexovacieho mechanizmu	8	Jakub Hlavačka	Done	19-02-22 10:37	17-03-22 12:31	AMS Sprint 6		
Story	Refaktoring MongoDB	2	Dominik Horvath	Done	19-02-22 10:37	17-03-22 12:31	AMS Sprint 6		
Story	Stránkovanie článkov na frontende	5	Táňa Poláková	Done	19-02-22 10:36	17-03-22 12:31	AMS Sprint 6		
Story	Stránkovanie API	13	Adam Šipka	Done	19-02-22 10:36	10-03-22 11:05	AMS Sprint 6		
Story	Autentifikácia usera	8	David Slady	Done	19-02-22 10:35	24-03-22 11:18	AMS Sprint 7		
Story	Komunikácia medzi aplikačným serverom a user databázou	2	David Slady	Done	19-02-22 10:35	21-03-22 19:32	AMS Sprint 7		
Story	Základná user tabuľka v user databáze	5	David Slady	Done	19-02-22 10:33	17-03-22 12:31	AMS Sprint 6		
Story	Password management		Dominik Horvath	Done	14-02-22 19:40	19-02-22 11:08			
Story	Zabezpečenie endpointov cez HTTP basic auth	3	Jakub Müller	Done	14-02-22 19:39	07-03-22 19:43	AMS Sprint 6		
Story	Verziovanie docker compose		David Slady	Done	14-02-22 19:35	24-02-22 12:48			
Story	Dokumentácia User guide		Táňa Poláková	Done	14-02-22 19:32				
Story	Dokumentácia elasticsearch		Jakub Hlavačka	Done	14-02-22 19:31				
Story	Dokumentácia fungovania NGINX		Adam Šipka	Done	14-02-22 19:28				
Story	Dokumentácia fungovania dockeru		Dominik Horvath	Done	14-02-22 19:28				
Story	Dokumentácia node serveru		David Slady	Done	14-02-22 19:27				
Story	Mapa domény na stránku tímu		Jakub Müller	Done	14-02-22 19:27	24-02-22 12:48			
Story	MongoDB unit test	3	Adam Šipka	Done	07-12-21 13:45	14-12-21 19:24	AMS Sprint 5		

Story	Testovanie dostupnosti kontajnerov	2	Dominik Horvath	Done	07-12-21 13:37	15-12-21 17:14	AMS Sprint 5	
Story	Konfigurácia NGINX	1	David Slady	Done	07-12-21 13:36	15-12-21 17:14	AMS Sprint 5	
Story	[RP] Exporty uloh	0.5	Táňa Poláková	Done	07-12-21 13:33	15-12-21 17:12	AMS Sprint 5	
Story	[RP] Globálna retrospektíva	1	Táňa Poláková	Done	07-12-21 13:31	15-12-21 17:09	AMS Sprint 5	
Story	[RP] Šprint č. 5	0.5	Táňa Poláková	Done	07-12-21 13:30	15-12-21 17:09	AMS Sprint 5	
Story	[RP] Metodika docker	1	Adam Šipka	Done	07-12-21 13:30	15-12-21 16:49	AMS Sprint 5	
Story	[RP] Metodika github actions	1	Adam Šipka	Done	07-12-21 13:30	15-12-21 17:14	AMS Sprint 5	
Story	[ID] Sekv. diagram celkového fungovania	1	Jakub Hlavačka	Done	07-12-21 13:27	08-12-21 11:04	AMS Sprint 5	
Story	[ID] Modul aplikačný server + testovanie	5	David Slady	Done	07-12-21 13:27	15-12-21 17:14	AMS Sprint 5	
Story	[ID] Modul frontend + testovanie	5	Táňa Poláková	Done	07-12-21 13:26	15-12-21 14:57	AMS Sprint 5	
Story	[ID] Modul na indexovanie + testovanie	5	Jakub Hlavačka	Done	07-12-21 13:25	10-12-21 16:05	AMS Sprint 5	
Story	[ID] Modul API server + testovanie	5	Jakub Müller	Done	07-12-21 13:25	14-12-21 18:42	AMS Sprint 5	
Story	[ID] Modul scraper + testovanie	5	Dominik Horvath	Done	07-12-21 13:20	13-12-21 18:22	AMS Sprint 5	
Story	Rozebrať PostgreSQL	3	Jakub Hlavačka	Done	23-11-21 14:50	28-11-21 14:48	AMS Sprint 4	
Story	Komunikácia medzi aplikačným serverom a API serverom	5	David Slady	Done	23-11-21 14:46	07-12-21 12:32	AMS Sprint 4	
Story	Komunikácia medzi API serverom a Elastic Search	13	Jakub Müller	Done	23-11-21 14:45	06-12-21 19:33	AMS Sprint 4	
Story	Zobrazenie výsledkov vyhľadávania	13	Táňa Poláková	Done	23-11-21 14:43	05-12-21 17:11	AMS Sprint 4	
Story	Komunikácia medzi MongoDB a API serverom	13	Adam Šipka	Done	23-11-21 14:42	07-12-21 12:23	AMS Sprint 4	
Story	Indexovanie stiahnutých dát	8	Jakub Hlavačka	Done	23-11-21 14:41	29-11-21 16:55	AMS Sprint 4	
Story	Fuzzy search			To Do	23-11-21 14:40			
Story	Dokumentácia k riadeniu projektu		Táňa Poláková	Done	23-11-21 9:00	23-11-21 20:39	AMS Sprint 3	
Story	Dokumentácia k inžinierskemu dielu		Jakub Hlavačka	Done	23-11-21 9:00	23-11-21 19:32	AMS Sprint 3	
Subtask	Vytvorenie articles db na virtuálnej masine		Jakub Hlavačka	Done	21-11-21 19:00	21-11-21 19:16	AMS Sprint 3	
Subtask	Analýza shard, nodes		Jakub Hlavačka	Done	16-11-21 12:59	21-11-21 18:49	AMS Sprint 3	
Subtask	Analýza River plugin		Jakub Hlavačka	Done	16-11-21 12:58	16-11-21 14:56	AMS Sprint 3	
Story	Rozšírenie docker compose o MongoDB a Elasticsearch	5	Jakub Hlavačka	Done	15-11-21 13:16	21-11-21 19:16	AMS Sprint 3	
Story	Deployment scraper	13	Dominik Horvath	Done	15-11-21 13:08	28-11-21 17:01	AMS Sprint 4	
Subtask	vytvoriť komprimovanu collection z main.py a zmerať čas indexovania		Jakub Hlavačka	Done	12-11-21 17:21	16-11-21 11:47	AMS Sprint 3	
Subtask	Vytvoriť config pre articles_index		Jakub Hlavačka	Done	12-11-21 17:21	12-11-21 17:21	AMS Sprint 3	
Subtask	Skonzultovať komentý v storke		Jakub Hlavačka	Done	12-11-21 16:12	15-11-21 13:04	AMS Sprint 3	
Subtask	Napísať dokumentáciu do README.md		Jakub Hlavačka	Done	12-11-21 15:46	16-11-21 11:44	AMS Sprint 3	
Subtask	Vytvorenie docker-compose file s MongoDB, Elasticsearch a Kibana		Jakub Hlavačka	Done	12-11-21 11:01	12-11-21 17:20	AMS Sprint 3	
Story	Implementácia dummy Flask API	5	Dominik Horvath	Done	09-11-21 14:45	15-11-21 19:43	AMS Sprint 3	

Story	Návrh Flask API		5	Táňa Poláková	Done	09-11-21 14:44	15-11-21 12:50	AMS Sprint 3	
Story	Analýza možností implementácie testovania		5	Adam Šipka	Done	09-11-21 14:43	23-11-21 20:39	AMS Sprint 3	
Story	Zabezpečenie komunikácie medzi klientom a serverom		8	David Slády	Done	09-11-21 14:38	23-11-21 14:36	AMS Sprint 3	
Story	Uprava dokumentácie zo stretnutí a šprintov tak, aby mohli ísť na stránku tímu		8	Táňa Poláková	Done	09-11-21 14:31	23-11-21 8:51	AMS Sprint 3	
Story	Integrácia MongoDB a Elasticsearch		8	Jakub Hlavačka	Done	09-11-21 14:22	16-11-21 12:54	AMS Sprint 3	
Story	Integrácia scraperu a vylepšenia parsera		5	Jakub Müller	Done	09-11-21 14:12	22-11-21 11:42	AMS Sprint 3	
Story	Rozšíriť úložisko pomocou nepriradeného disku		5	Dominik Horvath	Done	09-11-21 14:07	14-11-21 8:30	AMS Sprint 3	
Story	Zabrániť zacykleným buildom		5	Dominik Horvath	Done	07-11-21 23:07	15-11-21 8:22	AMS Sprint 3	
Story	PDF report z nájdených článkov		13	Jakub Hlavačka	Done	02-11-21 14:50	07-04-22 11:24	AMS Sprint 8	
Story	Príhásenie sa na odber nejakého vyhladaného človeka				To Do	02-11-21 14:50			
Story	Hodnotenie riziku pri nájdených článkoch				To Do	02-11-21 14:47			
Story	Zvýraznenie hľadanych slovíčok				To Do	02-11-21 14:45			
Subtask	geografická lokácia				Done	02-11-21 14:36	21-04-22 11:50	AMS Sprint 9	
Subtask	vyhladanie na základe zločínov				Done	02-11-21 14:35	21-04-22 11:50	AMS Sprint 9	
Subtask	obmedzenie vyhladávaní na základe dátumu vydania článkov				Done	02-11-21 14:33	21-04-22 11:50	AMS Sprint 9	
Story	Filter na web stránke		5	Táňa Poláková	Done	02-11-21 14:32	21-04-22 11:23	AMS Sprint 9	
Subtask	Iné možnosti prekladu článkov				To Do	02-11-21 14:23			
Subtask	Sekvenčné scrapovanie			Dominik Horvath	Done	02-11-21 14:09	06-11-21 14:10	AMS Sprint 04	
Story	Vyhľadanie RSS dvojičaťa k HTML článku - všeobecný postup			Dominik Horvath	Done	02-11-21 14:08	09-11-21 9:45	AMS Sprint 04	
Story	Technický návrh developerského prostredia		8	David Slády	Done	02-11-21 13:50	24-02-22 12:38	AMS Sprint 4	
Story	Napĺňanie MongoDB priamo pri scrapovaní			Adam Šipka	Done	02-11-21 13:44	14-11-21 14:06	AMS Sprint 04	AMS Sprint 3
Story	Rozšírenie google news scraperu na lokáciu UK			Dominik Horvath	Done	02-11-21 13:42	02-11-21 16:53	AMS Sprint 04	
Story	Analýza médií, ktoré dostaneme vo výstupoch a aké search queries treba poslať, aby sme mali výsledky ako z Google News				To Do	02-11-21 13:25			
Story	Pozrieť sa na zahodené články, kt. neobsahovali paragrafy				To Do	02-11-21 13:23			
Subtask	Ako mongoDB pracuje s vopred komprimovaným textom				Done	01-11-21 12:24	01-11-21 12:25	AMS Sprint 04	
Subtask	vyskusar nad dátach				Done	01-11-21 12:23	01-11-21 12:23	AMS Sprint 04	
Subtask	realne kniznice				Done	01-11-21 12:21	01-11-21 12:22	AMS Sprint 04	
Subtask	elastic search a komprimácia				Done	01-11-21 12:20	01-11-21 12:21	AMS Sprint 04	
Subtask	analýza vyhľadávania nad komprimovaným textom				Done	01-11-21 12:19	01-11-21 12:20	AMS Sprint 04	
Subtask	bezstratova kompresia text				Done	01-11-21 12:18	01-11-21 12:19	AMS Sprint 04	
Story	Pamätková zložitosť elastic search na vzorke dát			Jakub Hlavačka	Done	01-11-21 12:15	12-11-21 17:09	AMS Sprint 04	AMS Sprint 3
Story	Parsovanie tagov pomocou lxml, readability, traflitura			Táňa Poláková	Done	30-10-21 11:41	09-11-21 13:12	AMS Sprint 04	

Story	Zavesenie klienta	13	David Slady	Done	27-10-21 14:45	08-11-21 0:51	AMS Sprint 04	
Subtask	NGINX konfiguracia			Done	27-10-21 14:44	31-10-21 20:21	AMS Sprint 04	
Subtask	Kontajner Express server			Done	27-10-21 14:44	31-10-21 20:21	AMS Sprint 04	
Subtask	Kontajner Flask API			Done	27-10-21 14:44	30-10-21 12:53	AMS Sprint 04	
Story	Rozšíriť docker compose	13	Dominik Horvath	Done	27-10-21 14:43	31-10-21 20:22	AMS Sprint 04	
Story	Pridať info o projekte a opísať členov tímu	4	Jakub Müller	Done	27-10-21 14:42	01-11-21 11:37	AMS Sprint 04	
Story	Request URL s odpoveďou inou ako 200 uložiť do súboru	3	Adam Šipka	Done	27-10-21 14:40	01-11-21 11:35	AMS Sprint 04	
Story	Lokálna MongoDB so získanými dátami	8	Adam Šipka	Done	27-10-21 14:35	01-11-21 11:35	AMS Sprint 04	
Story	Analýza bezstratovej kompresie textu	8	Jakub Hlavačka	Done	27-10-21 14:25	01-11-21 12:25	AMS Sprint 04	
Story	Parsovanie tagov pomocou regex	8	Táňa Poláková	Done	27-10-21 14:19	02-11-21 13:49	AMS Sprint 04	
Story	Parsovanie tagov pomocou CSS Selector	5	Jakub Müller	Done	27-10-21 14:17	08-11-21 19:48	AMS Sprint 04	
Story	Stránka tímu		Jakub Müller	Done	18-10-21 15:03	01-11-21 11:37		
Story	Otvorenie portov na serveri		David Slady	Done	18-10-21 15:01	18-10-21 17:05		
Story	Porovnanie výsledkov nášho vyhľadávania s výsledkami Google News			To Do	13-10-21 14:54			
Story	Vyriešiť, čo s článkami, ktoré budú v pôvodnom zdroji nekompletné			To Do	13-10-21 14:47			
Story	Vyriešiť, aby filter nemožno urobiť "sudo su" na virtuálnom stroji (aby nemal šancu sa zmeniť na roota)		Táňa Poláková	Done	13-10-21 14:34	19-10-21 18:04		
Subtask	Zistiť, ako funguje Google Translate API, ako ho vieme použiť		Adam Šipka	To Do	13-10-21 14:32			
Story	Zistiť, ako fungujú iné zdroje Google News, prípadne aký mechanizmus používa Google News.		Adam Šipka	To Do	13-10-21 14:31			
Story	User databáza + zberanie aktivity používateľa	5	David Slady	Done	13-10-21 14:29			
Story	Prieskum databáz		Jakub Müller	Done	13-10-21 14:26	18-10-21 14:28		
Story	Vytvoríť github projekt		Jakub Müller	Done	13-10-21 14:22	18-10-21 12:09		
Story	Zistiť, čo spája kriminálne články			To Do	13-10-21 14:13			
Story	Príháška na TP CUP			Done	13-10-21 14:13	02-11-21 12:50		
Story	Set up elastic search v docker container		Jakub Hlavačka	Done	13-10-21 14:12	16-10-21 16:45		
Subtask	Redukcia a kategorizácia zoznamu zločinov		Dominik Horvath	Done	13-10-21 14:10	23-11-21 11:24		
Story	Získanie dát pre prototyp		Dominik Horvath	Done	13-10-21 14:09	26-10-21 13:29		
Story	Vytvoríť Slack server		Táňa Poláková	Done	05-10-21 13:49	11-10-21 12:08		
Story	Analýza parametrov Google News		Adam Šipka	Done	05-10-21 13:43	13-10-21 14:26		
Story	Analýza - multilingvistika			To Do	05-10-21 13:37			
Subtask	API / scrapovanie		Dominik Horvath	Done	05-10-21 13:32	10-10-21 12:07		
Subtask	Získanie vzorky RSS		Jakub Hlavačka	Done	05-10-21 13:31	08-10-21 20:27		

Story	Realny scrapping - vytvorenie vzorky		Dominik Horvath	Done	05-10-21 13:27	10-10-21 18:14		
Story	Zistit, ktoré miestnosti su na karticky		David Slady	Done	05-10-21 13:24	18-10-21 16:12		
Story	Nasu API dat na portal verejnych API v ramci propagacie			To Do	05-10-21 13:22			
Story	Lofi pouzivatel'skeho rozhrania - main page		David Slady	Done	05-10-21 13:20	20-02-22 10:43		
Story	Revizia poziadaviek		Táňa Poláková	Done	05-10-21 13:19	13-10-21 14:05		
Story	Spísať denník z prvého stretnutia s vedúcim		Táňa Poláková	Done	03-10-21 15:50	05-10-21 13:30		
Story	Pripady pouzitia		Dominik Horvath	Done	01-10-21 14:23	05-10-21 12:08		
Story	Prieskum získavania dát, práca s knižnicou		Adam Šipka	Done	01-10-21 14:22	05-10-21 13:24		
Story	Specifikacia požiadaviek		Jakub Havačka	Done	01-10-21 14:22	03-10-21 15:17		
Story	Vysoka architektura		David Slady	Done	01-10-21 14:21	03-10-21 14:56		
Story	Zaobstarat stroj v škole		Táňa Poláková	Done	01-10-21 14:18	13-10-21 14:23		
Story	Zoznam trestných činov, ktoré budu použite ako queries na google news		Jakub Müller	Done	01-10-21 14:18	04-10-21 21:20		