

Tímový projekt  
Projektová dokumentácia

Míľnik č. 2

**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

Ak. Rok: 2021/2022

## Inžinierske dielo

Míľnik č. 2

**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

Ak. Rok: 2021/2022

<b>Big Picture</b>	<b>4</b>
Úvod	4
Globálne ciele projektu na zimný semester	4
Vzorka dát	4
Čistenie dát	4
Uloženie dát	4
Webová stránka tímu	5
Indexovanie dát	5
Funkčné webové rozhranie pre AMS	5
Funkčná API	5
Prototyp so statickým sťahovaním a indexovaním	5
Celkový pohľad na systém	6
Vysoká architektúra	6
Diagram prípadu použitia	7
Sekvenčný diagram vyhľadávania v systéme	8
Sekvenčný diagram indexovania	9
Sekvenčný diagram fungovania webovej aplikácie	10
Sekvenčný diagram fungovania API serveru	11
Sekvenčný diagram fungovania scrapovania	12
<b>Moduly systému</b>	<b>13</b>
Modul Elasticsearch	13
Analýza	13
Návrh	13
Implementácia	14
Testovanie	14
Modul Indexer	15
Analýza	15
Návrh	15
Implementácia	16
Testovanie	16
Modul Webová Aplikácia	17
Analýza	17
Návrh	17
Implementácia	18
Testovanie	19
Modul API server	20
Analýza	20
Návrh	20
Implementácia	20
Testovanie	21
Modul Scraper	22
Analýza	22
Návrh	22
Implementácia	22

Testovanie	23
<b>Prílohy</b>	<b>23</b>
Modul na indexer	24
Modul webovej aplikácie	24
Modul na server API	24
Modul na scraper	24

# 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 pre tento systém počas zimného semestra. Ďalšia kapitola je venovaná celkovému pohľadu na systém, ktorý sme priblížili 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 naplňali.

## Vzorka dát

Dôležitou časťou pred začiatkom implementácie projektov, ktoré pracujú s veľkým množstvom dát je potreba tieto dáta spoznať.

Náplňou tohto cieľa bolo vymyslieť spôsob, akým budeme získavať potrebné dáta pre náš systém, následne zistiť ich štruktúru, veľkosť a podobne.

Vzorkou dát boli články o zločinoch, ktoré sme sťahovali pomocou Google RSS. Veľkosť vzorky bola ~30GB, bolo v nej zahrnutých cez 100 zločinov, časová perióda bola jeden mesiac, samotné články pochádzali z regiónu USA a štruktúra článkov bola v HTML formáte.

## 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 je preň text z nadpisov a paragrafov článku absolútne postačujúci.

Ďalej sme sa aplikovali kompresiu Huffmanovho kódovania na ešte menšie vyťaženie úložného priestoru našich dát.

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 SSH kľúča na prístup do tímového virtuálneho stroja sa hlavným cieľom stalo vytvoriť webovú stránku. Táto webová stránka má umožňovať jednoduchý prístup k zázpisnicam zo stretnutí, exportom úloh a ostatných dokumentov. Jej ďalšou úlohou je predstavenie tímu č. 8.

## Indexovanie dát

Hlavnou funkcionalitou systému Adverse Media Screening má byť možnosť vyhľadávania fyzických a právnických osôb spojených so zločinmi, ktoré údajne vykonali. Aby sme tieto osoby a zločiny dokázali jednoducho vyhľadávať, je potrebné indexovať obsah stiahnutých článkov.

## Funkčné webové rozhranie pre AMS

Predpokladáme, že používatelia, ktorí budú náš system používať, nemajú znalosti ako jeho vývojári alebo iní kolegovia z IT oblasti. Z tohto dôvodu bolo úlohou ďalšieho cieľa vytvoriť užívateľsky priateľské rozhranie, ktoré je používateľom ľahko prístupné vďaka svojej webovej podstate.

## 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 funkcionalitu napríklad vo svojom vlastnom systéme.

## Prototyp so statickým sťahovaním a indexovaním

Výsledkom tohto globálneho cieľa je prvý funkčný prototyp, ktorý používateľovi umožňuje vyhľadávať nad aktuálnou databázou článkov. Databáza obsahuje ~51000 článkov z obdobia 15.11.2021 - 29.11.2021. Pre jednoduché, príjemné a intuitívne vyhľadávanie je používateľovi sprístupnené grafické rozhranie na stránke:

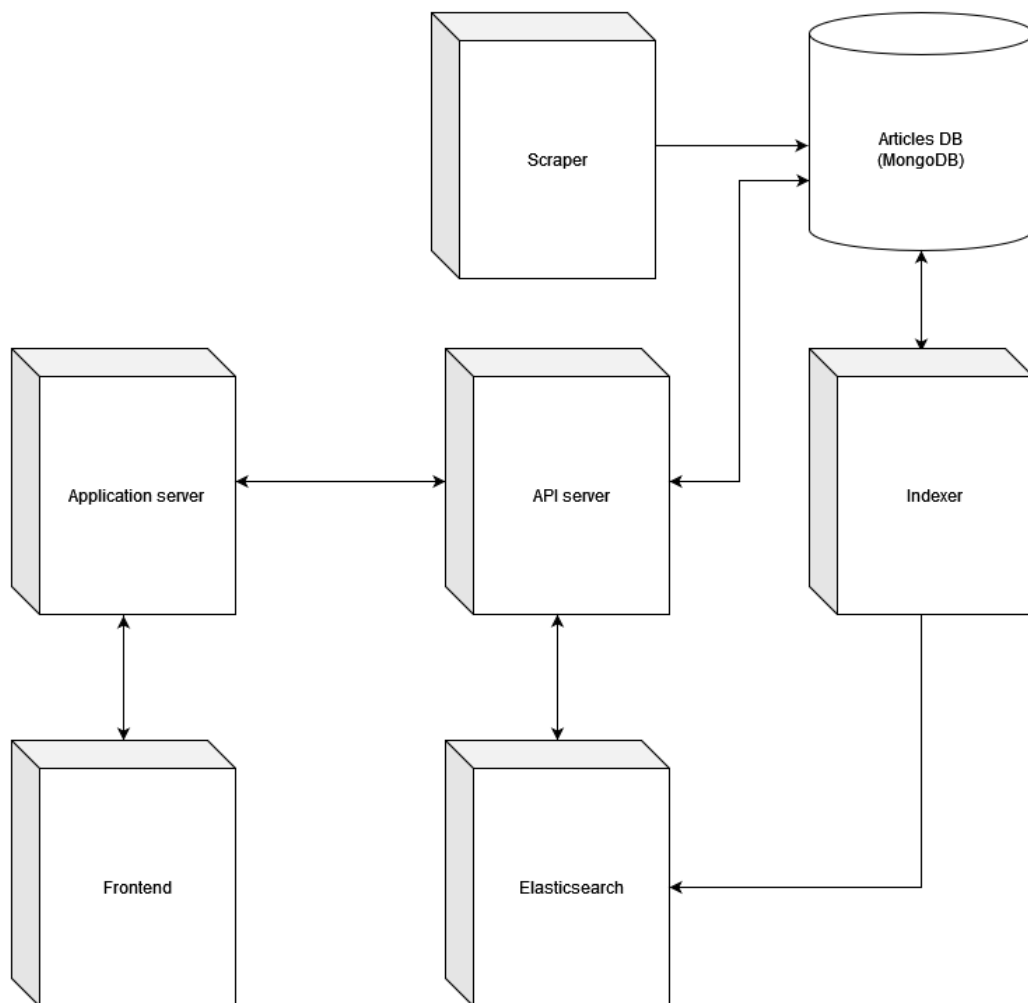
<https://team08-21.studenti.fiit.stuba.sk/ams/>

## 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.

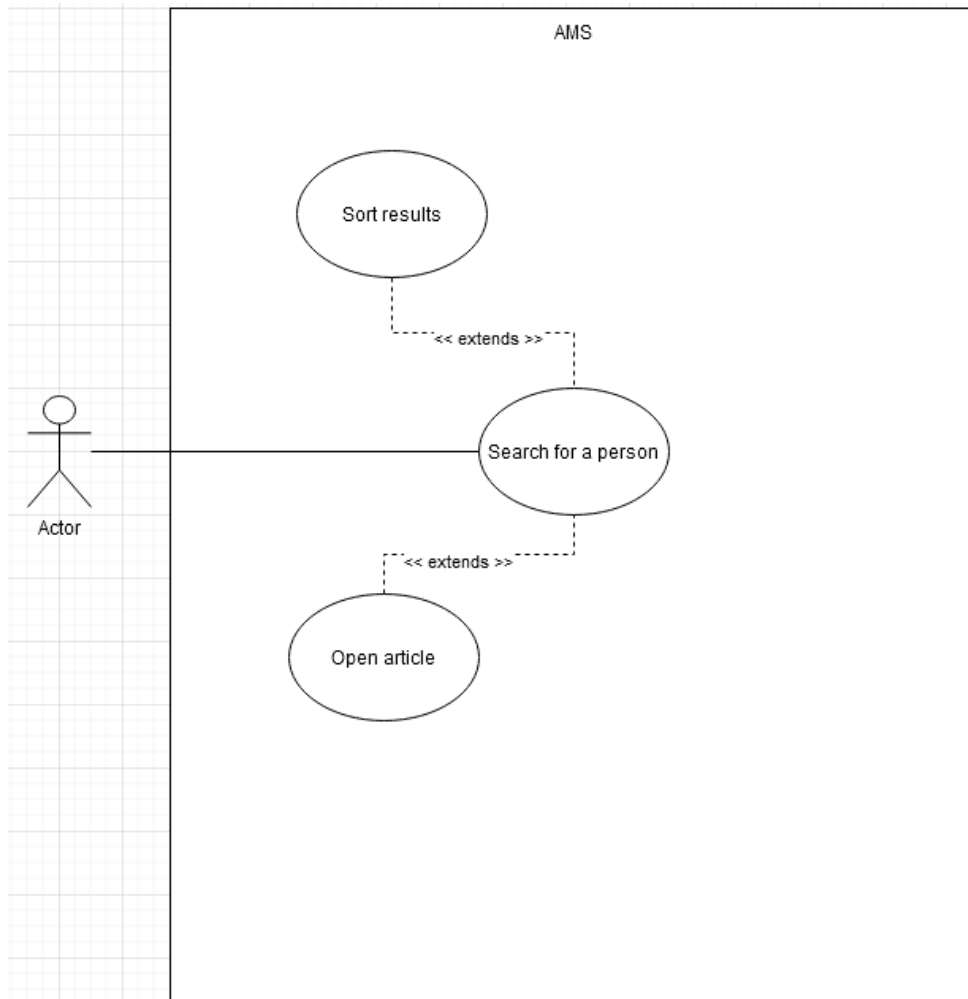
### Vysoká architektúra

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



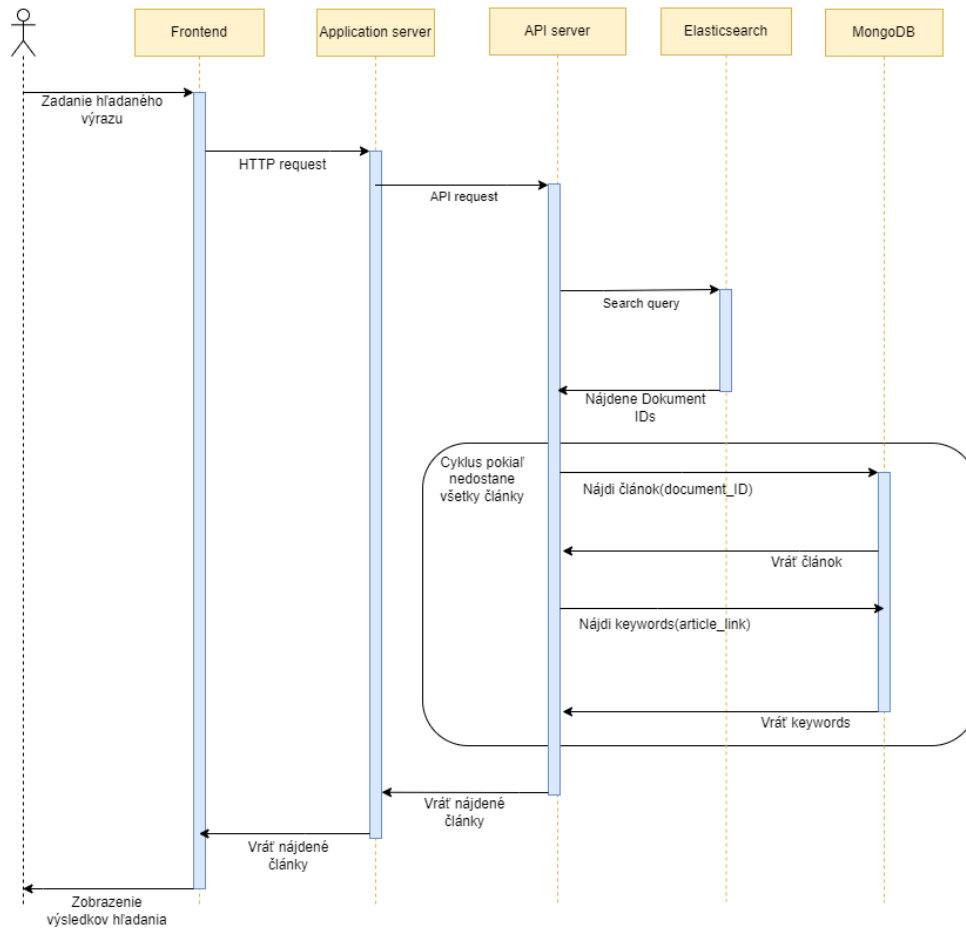
## Diagram prípadu použitia

Diagram je jednoduchý vzhľadom na jednoduchosť nášho systému. Funkcionalita a prípady použitia sa budú častom rozvíjať.

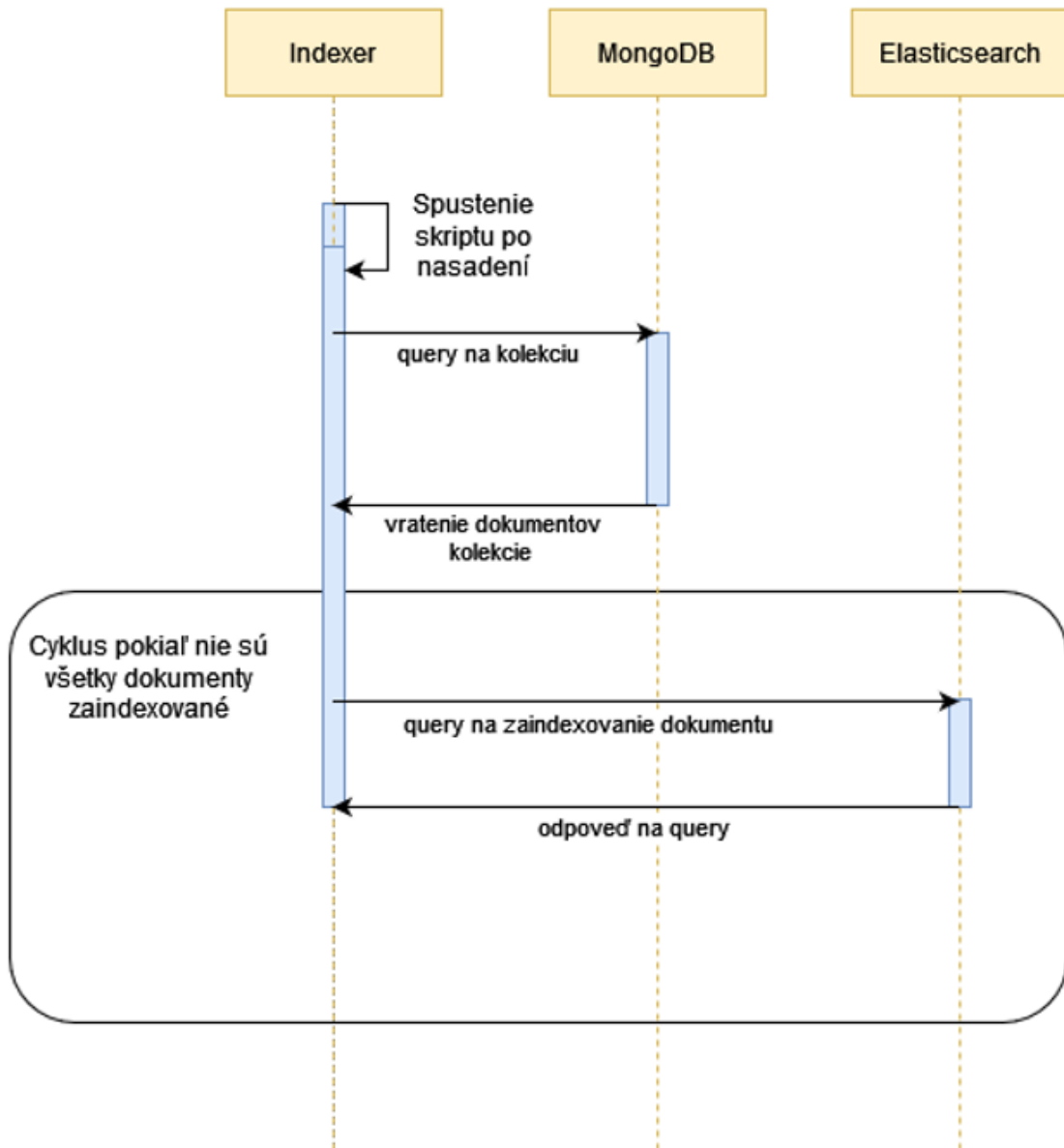




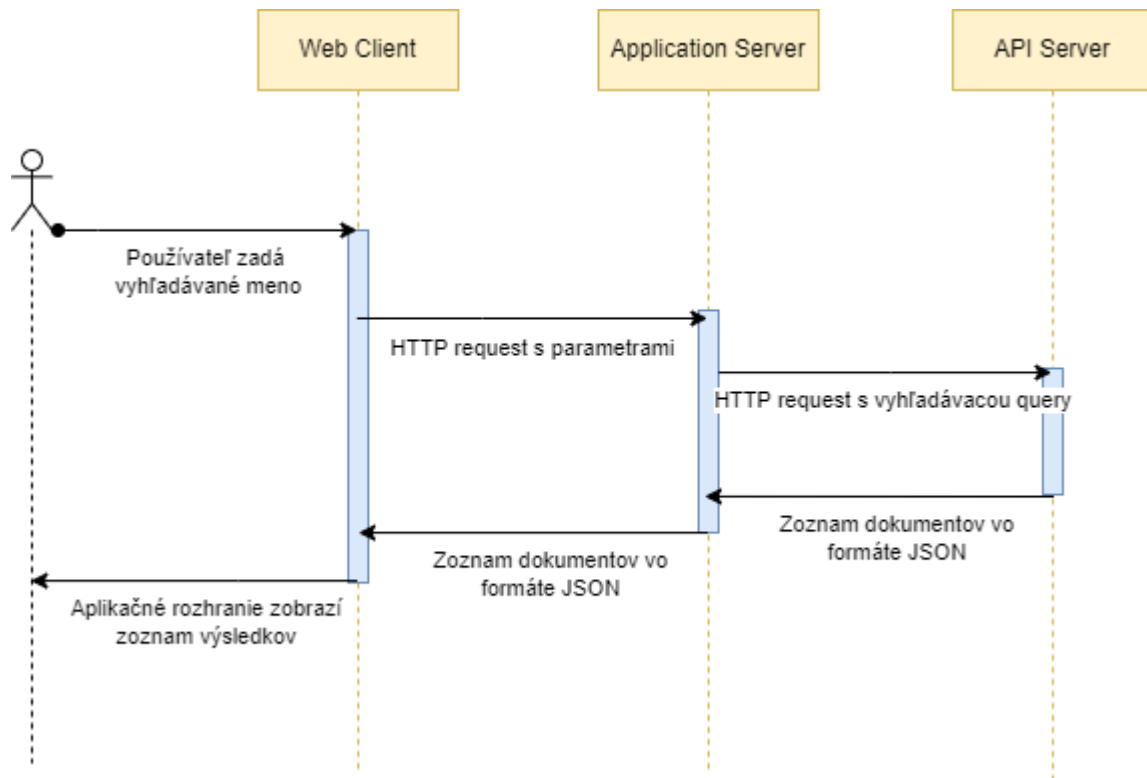
## Sekvenčný diagram vyhľadávania v systéme



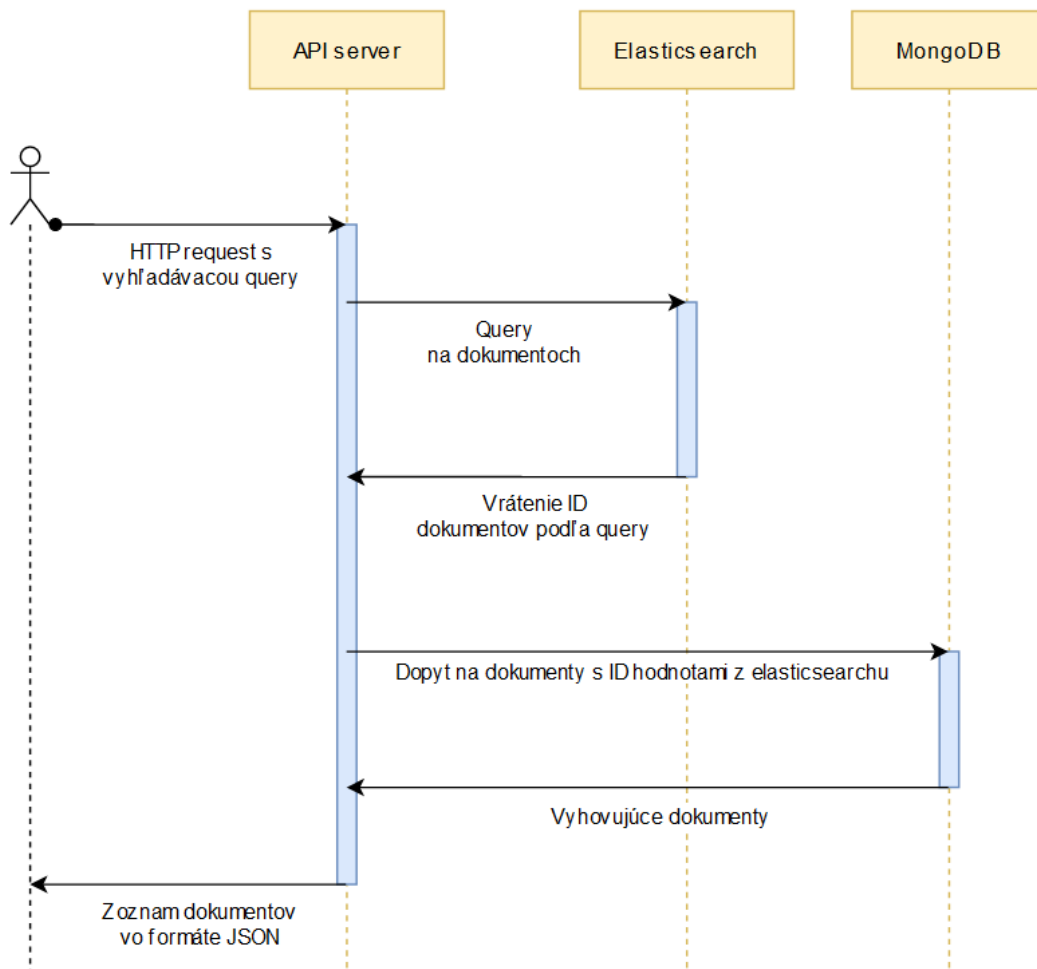
## Sekvenčný diagram indexovania



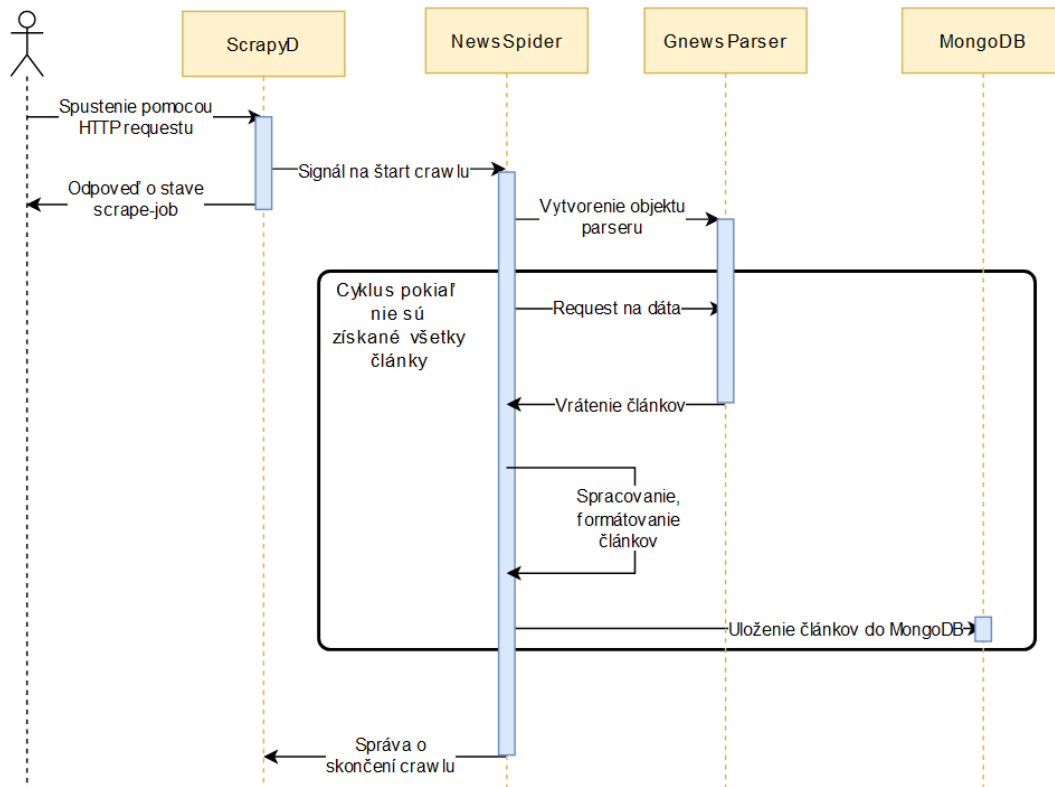
## Sekvenčný diagram fungovania webovej aplikácie



## Sekvenčný diagram fungovania API serveru



## Sekvenčný diagram fungovania scrapovania



# Moduly systému

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

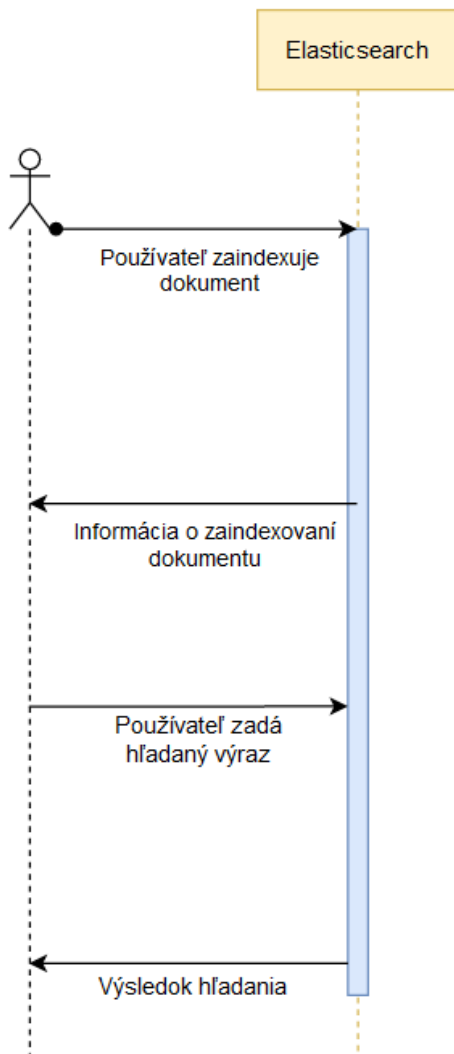
## Modul Elasticsearch

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 fungujeme na jednom uzle, ktorý má jeden shard a jednu repliku. V samotnom sharde a aj v replike je uložený Lucene index.

Elasticsearch využíva pri indexovaní a vyhľadávaní štandardný analyzátor s filtrovaním anglických stop slov.

Každý dokument je zložený z jedného textového poľa (angl. field), ktorého hodnota je obsah jedného článku z databázy.

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.

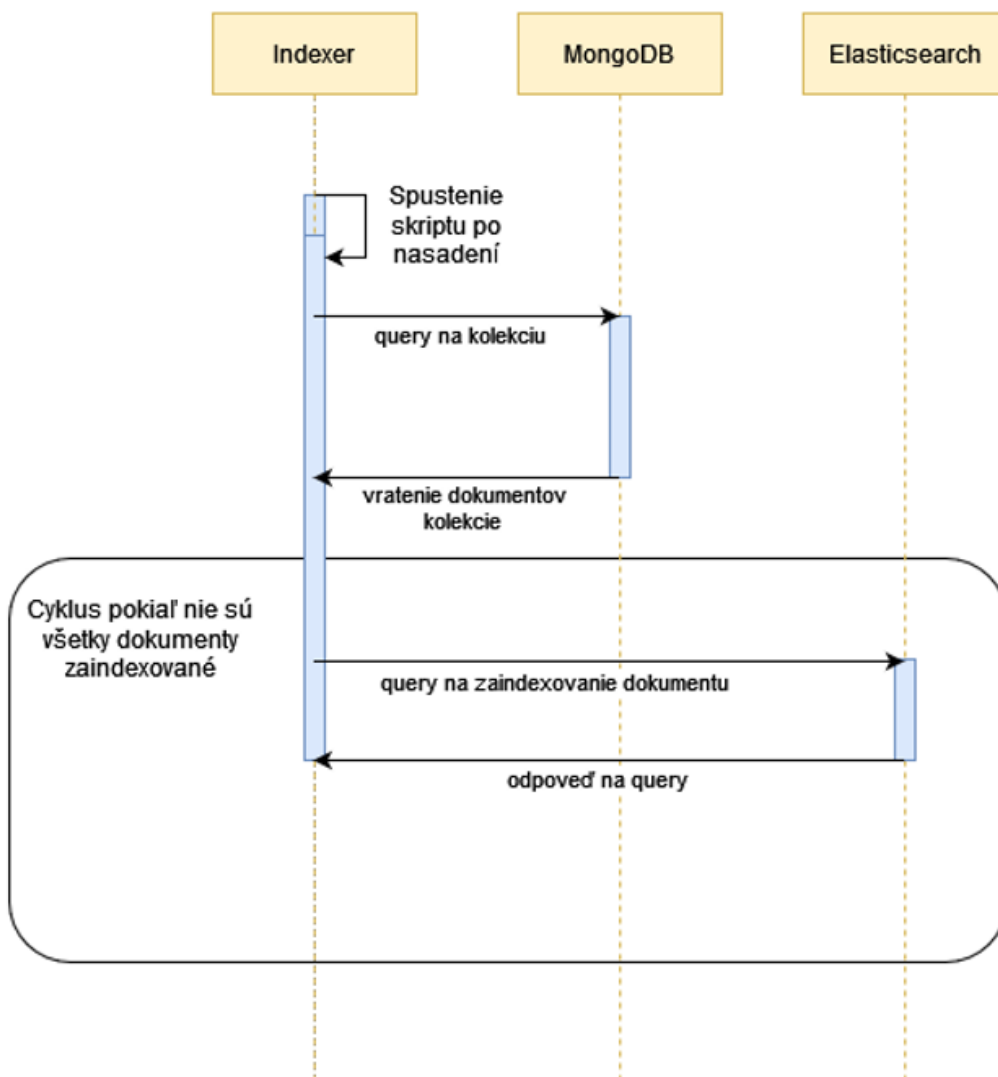
## Modul Indexer

V tejto sekcii sa bližšie pozrieme na modul systému, ktorý súvisí s indexovaním dát.

### Analýza

Samotné dáta boli stiahnuté prostredníctvom modulu Scraper, ktorý ich aj očistil a uložil do distribuovanej Mongo databázy. Aktuálne fungujeme na statickom aktualizovaní indexu, ktoré sa spustí po nasadení kontajneru indexer\_service. Úlohou skriptu v kontajneri je vytiahnuť dáta z úložiska a poslať ich nástroju Elasticsearch, ktorý vykoná potrebné indexovanie.

### Návrh





## Implementácia

Python script na indexovanie stiahnutých článkov v MongoDB je nasadený prostredníctvom technológie Docker. Script, ktorý vykoná indexovanie dokumentov sa automaticky spustí po nasadení Docker kontajnera.

MongoDB je nasadená na virtuálnom stroji prostredníctvom technológie Docker. V databáze admin sa nachádza kolekcia articles, ktorej obsah je komprimovaný Huffmanovým kódovaním za účelom šetrenia úložiska.

Jeden element kolekcie obsahuje id, title (názov článku), published (dátum vydania článku), link (link na článok) a html (obsah článku) pole.

Script využíva knižnice ako Pymongo, prostredníctvom ktorej sa pripája ku Mongo databáze. Ďalšou dôležitou knižnicou je Elasticsearch, vďaka ktorej dokážeme poslať údaje samotnému nástroju Elasticsearch, ktorý ich následne indexuje.

## Testovanie

Elasticsearch a MongoDB boli testované prostredníctvom jednotkových textov.

Celá služba bola otestovaná ručne nasledovným spôsobom. Script vytiahol údaje z databázy, následne ich zaindexoval v Elasticsearch a potom sme správne indexovanie overovali ručne prostredníctvom nástroja Kibana.

## Modul Webová Aplikácia

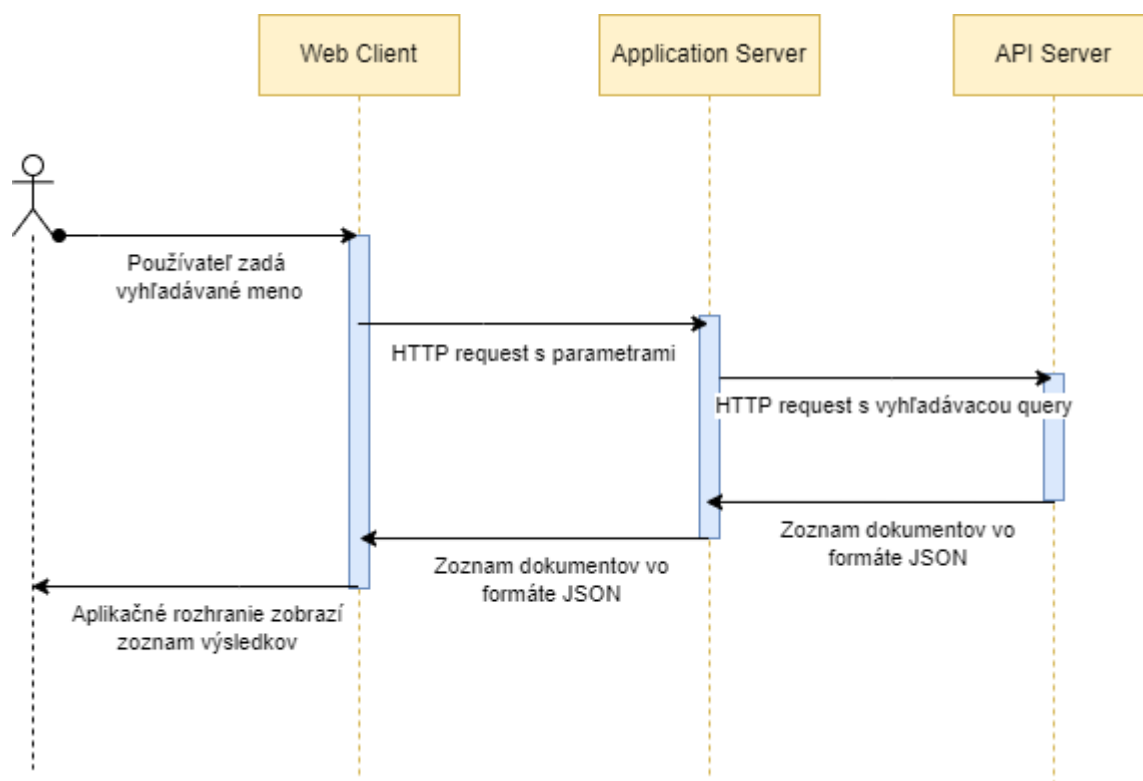
V tejto sekcii sa bližšie pozrieme na aplikačnú časť nášho riešenia.

### Analýza

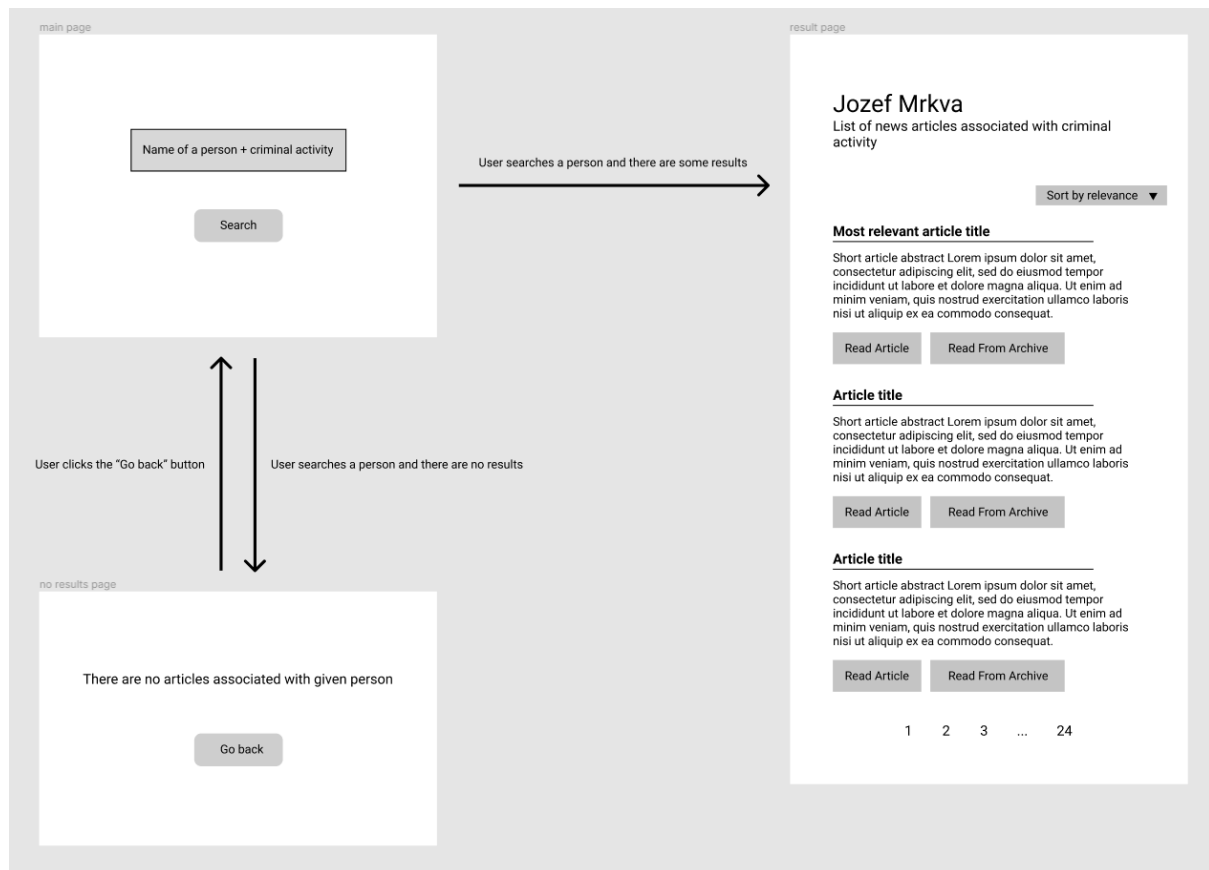
Webová aplikácia je potrebná pre zabezpečenie užívateľského rozhrania pre jednoduchší prístup na prácu s dátami. Aplikačná časť je tvorená webovým klientom a aplikačným serverom. Aplikačný server zabezpečuje komunikáciu s ostatnými modulmi. Webový klient slúži používateľovi ako grafické rozhranie.

### Návrh

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



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.



## Implementácia

Implementácia je zabezpečená pomocou klient aplikácie napísanej v ReactJS a NodeJS aplikačného servera, ktorý funguje pomocou knižnice ExpressJS.

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.

Klient poskytuje základné zobrazenie výsledkov vyhľadávania a to v dvoch stavoch. Prvým je domovská stránka a druhým je zobrazenie výsledkov. Na domovskej stránke je možné zadať požadovaný reťazec, ktorý je poslaný na server. Odpoveď zo servera je následne zobrazená na "druhej" stránke (aplikáciu sme implementovali ako single page) v podobe zoznamu názvov, odkazov, dátumov publikovania a kľúčových slov pre nájdené články. Používateľa informujeme, koľko výsledkov bolo vrátených na požadovaný dopyt, takže sme neimplementovali ďalšiu stránku v prípade žiadnych výsledkov.

Zatiaľ sme neimplementovali žiadne prípady použitia, ktoré by vyžadovali aplikačnú databázu, plánujeme však použiť PostgreSQL pre zabezpečenie možnosti prihlasovania a rozšírenej funkcionality pre prihlásených používateľov.

Nasledujúci obrázok predstavuje zobrazenie výsledkov po zadaní dopytu "john".

ams

Search

john

SUBMIT

10 results found.

25 Nov 2021 | [https://www.contactmusic.com/elton-john/news/sir-elton-john-announces-hometown-shows-at-watford-fc-s-vicarage-road\\_6337720](https://www.contactmusic.com/elton-john/news/sir-elton-john-announces-hometown-shows-at-watford-fc-s-vicarage-road_6337720)

[Sir Elton John announces hometown shows at Watford FC's Vicarage Road - Contactmusic.com](https://www.contactmusic.com/elton-john/news/sir-elton-john-announces-hometown-shows-at-watford-fc-s-vicarage-road_6337720)

Treason

23 Nov 2021 | <https://www.wsj.com/podcasts/bad-bets/enron-ep-7-the-trial/709cac07-5524-4d01-8959-32daaaa11299>

[Enron, Ep 7: The Trial - Bad Bets - WSJ Podcasts - The Wall Street Journal](https://www.wsj.com/podcasts/bad-bets/enron-ep-7-the-trial/709cac07-5524-4d01-8959-32daaaa11299)

Bankruptcy Fraud

16 Nov 2021 | <https://www.wsj.com/podcasts/bad-bets/enron-ep-6-lies-and-choices/3dfac6b4-4307-4970-9fe8-403cfc8d4410>

[Enron, Ep 6: Lies and Choices - Bad Bets - WSJ Podcasts - The Wall Street Journal](https://www.wsj.com/podcasts/bad-bets/enron-ep-6-lies-and-choices/3dfac6b4-4307-4970-9fe8-403cfc8d4410)

Tax Evasion

16 Nov 2021 | <https://screenrant.com/netflix-action-movies-john-wick-copycats-similarities-explained/>

[Why Netflix Is Obsessed With Ripping Off John Wick | Screen Rant - Screen Rant](https://screenrant.com/netflix-action-movies-john-wick-copycats-similarities-explained/)

Assassination

17 Nov 2021 | <https://bluebonnetnews.com/2021/11/17/houston-fbi-seeking-info-on-child-sexual-assault-victim/>

[Houston FBI seeking info on child sexual assault victim - Bluebonnet News](https://bluebonnetnews.com/2021/11/17/houston-fbi-seeking-info-on-child-sexual-assault-victim/)

Child Exploitation

17 Nov 2021 | <https://www.campaignlive.co.uk/article/john-lewis-legal-fight-2019-christmas-ad/1733343>

[John Lewis in legal fight over 2019 Christmas ad - CampaignLive](https://www.campaignlive.co.uk/article/john-lewis-legal-fight-2019-christmas-ad/1733343)

Copyright Matters

19 Nov 2021 | <https://www.cbc.ca/news/canada/saskatoon/drunk-driver-back-behind-bars-1.6255127>

[Man who killed Saskatoon teen in 2013 drunk driving crash back behind bars for driving while disqualified - ...](https://www.cbc.ca/news/canada/saskatoon/drunk-driver-back-behind-bars-1.6255127)

Drunk Driving

17 Nov 2021 | <https://jezebel.com/john-mayer-is-a-dick-but-he-doesn-t-deserve-death-thre-1848099884>

[John Mayer Is a Dick, But He Doesn't Deserve Death Threats - Jezebel](https://jezebel.com/john-mayer-is-a-dick-but-he-doesn-t-deserve-death-thre-1848099884)

Death Threat

<< < 1 2 > >>

## Testovanie

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

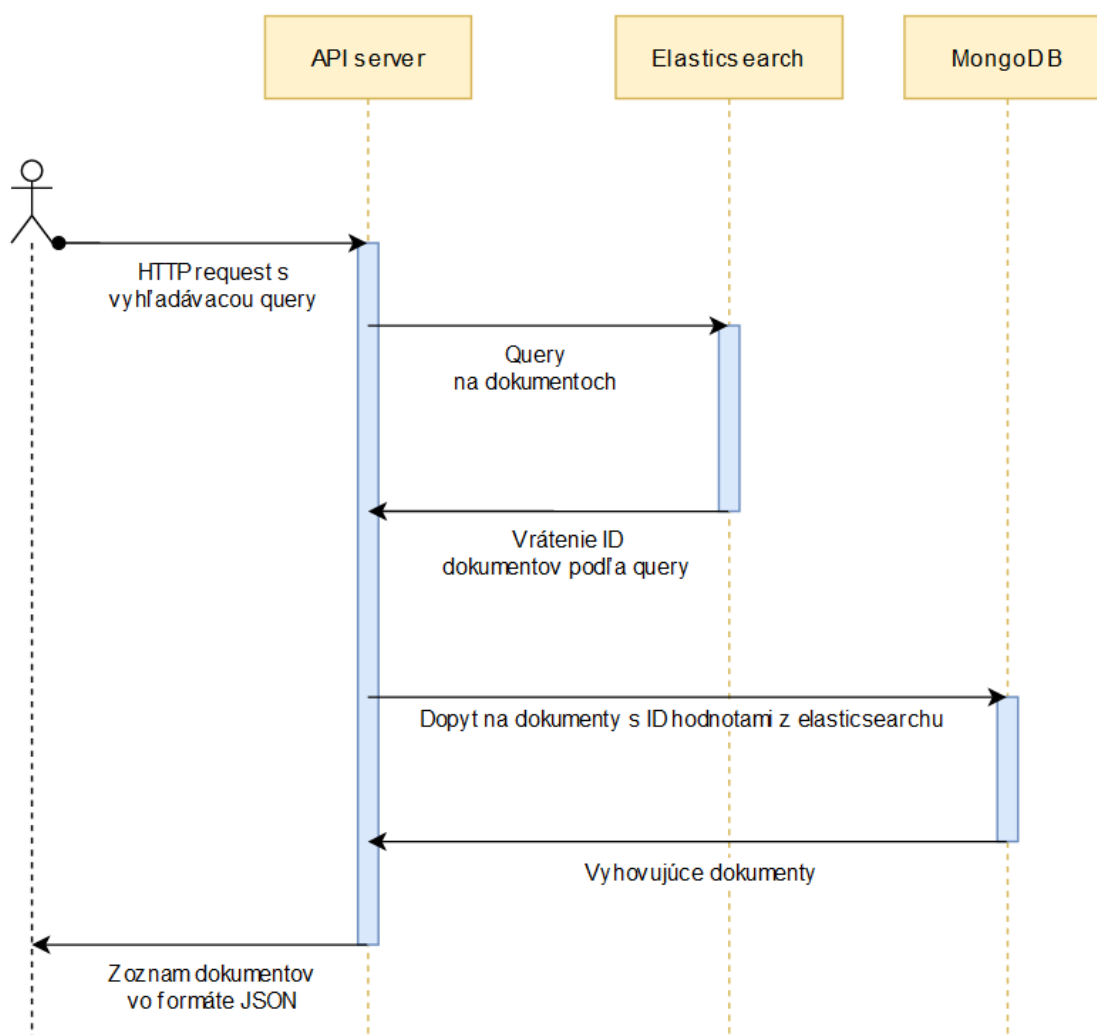
## Modul API server

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`, ktorá akceptuje jeden povinný parameter `q`, ktorý

obsahuje vyhľadávaný reťazec. Ďalšie parametre sú smerované na redukciu vyhľadávania, ako napríklad časové rozpätie alebo lokalizácia (jazyk a krajina pôvodu). Prvá verzia API však slúžila iba na testovanie.

Na reálne použitie bola vytvorená druhá verzia API, ktorá bola implementovaná v prototypu nášho projektu. Táto verzia API sa dokáže 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.

## Testovanie

API sa 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. Okrem tohto sme testovali, či správne funguje vyťahovanie kľúčových slov z kolekcie *crimemaps*.

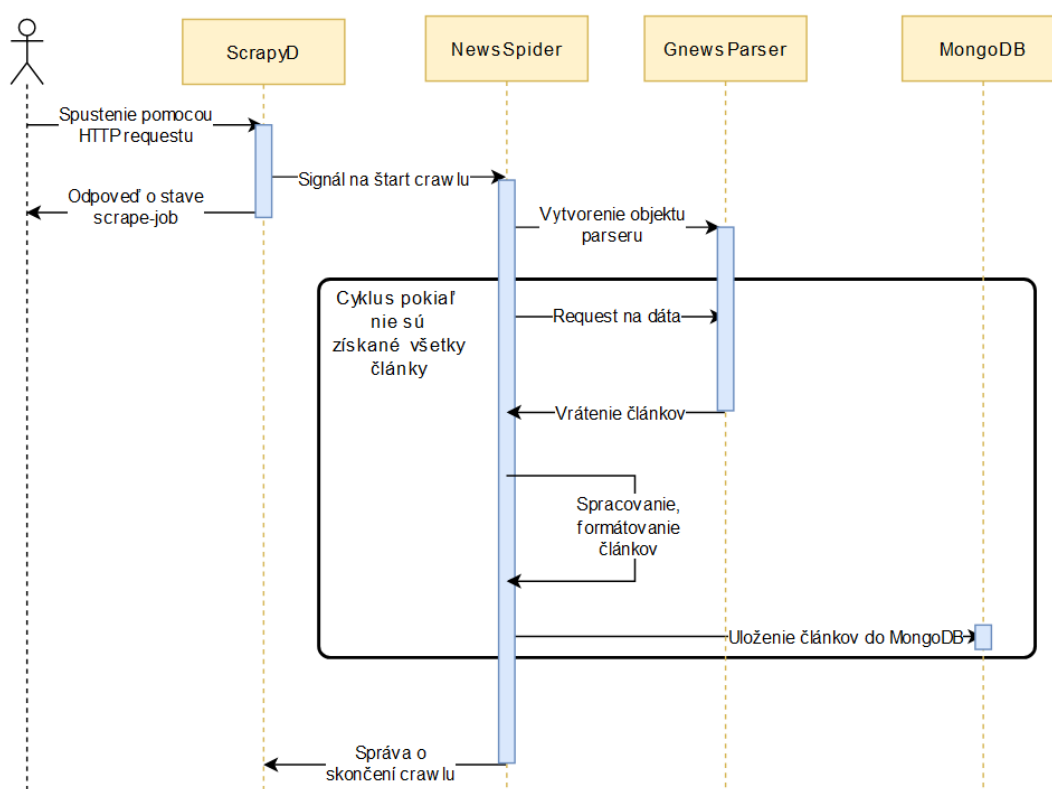
## Modul Scraper

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, aby nestratili svoju výpovednú hodnotu. Nakoniec naše dáta scraper ukladá do Mongo databázy.

### 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.

## 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 porovnala 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 sa testuje 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

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

### Modul na indexer

Návod na spustenie tohto modulu je opísaný v README.md tohto [repozitáru](#).

### Modul webovej aplikácie

Návod na spustenie frontendovej a backedovej časti modulu Webovej aplikácie je v README tohto [repozitáru](#).

### Modul na server API

Krátka dokumentácia k jednotlivým HTTP requestom je dostupná v README [repozitáru](#). Obsahuje taktiež Dockerfile pre vytvorenie Docker image na testovanie API na lokálnom prostredí.

### Modul na scraper

Návod na spustenie scraperu sa nachádza v README [repozitáru](#). Je tu taktiež návod, ktorým sa spustia jednotkové testy modulu.

## Riadenie projektu

Míľnik č. 2

**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

<b>Big Picture</b>	<b>4</b>
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	11
Šprint č. 2	13
Šprint č. 3	15
Šprint č. 4	17
Šprint č. 5	19
Globálna retrospektíva za zimný semester	21
<b>Motivačný dokument</b>	<b>22</b>
Tím	22
Motivácia	23
Téma č. 4 - Adverse Media Screening [AMS]	23
Téma č. 8 - Educational Content Engineering Hub - Databáza otázok, odpovedí, úloh a riešení [ECEH-DU]	24
Príloha A - Zoradenie všetkých tém podľa priority	25
Príloha B - Rozvrh tímu	26
<b>Metodiky</b>	<b>28</b>
Metodika Scrum	28
Standup	28
Šprint review	28
Plánovanie	28
Práca v nástroji Jira	28
Backlog	29
Nástenka	29
Git Metodika	30
Vetvenie	30
Commit	30
Pull requests	30
Metodika nasadzovania	31
Docker	31
Komunikácia kontajnerov	31

Riadenie projektu

GitHub Actions	31
Build & Deploy	31
<b>Export evidencie úloh</b>	<b>33</b>

# Big Picture

## Úvod

Riadenie projektu je kľúčový aspekt, ktorý ovplyvňuje či bude projekt úspešne ukončený, alebo nie. Zo začiatku semestra sme tejto oblasti venovali málo pozornosti, čo sa čiastočne odzrkadlilo aj na našich výsledkoch. Človek sa najlepšie učí na vlastných chybách a my sme si dali záväzok neodsúvať náležitosti agilného vývoja bokom.

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 zimný semester.

## 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 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	Dáta, Dokumentácia	Big Picture - inžinierske dielo, Modul - Elasticsearch a indexovanie, Sekvenčný diagram vyhľadávania v systéme
Dominik Horváth	Manažér plánovania	Dáta, API server	Diagram prípadu použitia, Modul - API server, Modul - scraper
Jakub Müller	Code review, Manažér kvality	Stránka tímu, Dáta	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 - webová aplikácia
Dávid Silady	Hlavný architekt	Webová aplikácia	Modul - webová aplikácia
Adam Šípka	Manažér testovania	Dáta, Databáza článkov	Dokumentácia metodík

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

Tak, ako každý z nás, na začiatku vývoja prišiel Jakub do kontaktu hlavne s prieskumnými úlohami. Z dlhodobého hľadiska sa ale zamerával najmä na prácu s dátami, čo zahŕňalo implementovanie získavania dát a následne ich indexovanie. 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 modul Elasticsearch a indexovanie.

## Dominik Horváth

Dominik venoval takmer všetok svoj čas práci na projekte dátam. Na začiatku mu však vystala aj úloha urobiť diagram prípadov použitia, ktorý je veľmi jednoduchý vzhľadom na povahu nášho softvéru. Diagram je súčasťou dokumentácie inžinierskeho diela. Okrem toho zdokumentoval modul API server a modul Scraper. Spoločne s ostatnými členmi tímu sa podieľal na tom, že sme získali vzorku dát a na nej postavili systém vyhľadávania. Má zásluhu aj na tom, ako bude vyzeráť komunikácia medzi klientom a serverom z dátovej časti (API) a komunikáciu aj zdokumentoval. Dominikova rola je manažér plánovania.

## 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. 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, exporthy úloh upravila tak, aby bolo v nich ľahko vyhľadávateľné a taktiež upravila zápisy zo stretnutí,

ktoré sú na stránke tímu. V rámci modulov je spoluautorom dokumentácie modulu Webovej aplikácie.

### Dávid Silady

Dávidove skúsenosti s tvorbou webu sme využili pri našom vlastnom webe projektu. Svoj čas nevenoval takmer ničomu inému, než náležitostiam súvisiacim s webom. Medzi ne patrí aj návrh používateľského rozhrania, komunikácia medzi klientom a serverom a konfigurácia servera. Dávid je spoluautorom dokumentácie vysokej architektúry a aj preto sme ho zvolili ako hlavného architekta. Dokumentácia modulu Webová aplikácia je aj jeho zásluha.

### Adam Šípka

Aby sme vedeli, ako funguje knižnica, ktorá je naším zdrojom získavania údajov, Adam ju ako jednu z prvých úloh analyzoval. Neskôr sa začal venovať Mongo databáze, ktorú spravuje dodnes. Taktiež, časť implementácie získavania dát patrí práve Adamovi. Rola, ktorej sa zaviazal, že sa bude poctivo venovať a študovať potrebné náležitosti, je manažér testovania. Adam zdokumentoval 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, ohlási 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.

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

## Riadenie projektu

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. Druhý 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 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

<b>Meno</b>	<b>Súčet počtu odhadovaných bodov</b>	<b>Súčet počtu pridelených bodov</b>	<b>Podiel práce (%)</b>
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

<b>Meno</b>	<b>Súčet počtu odhadovaných bodov</b>	<b>Súčet počtu pridelených bodov</b>	<b>Podiel práce (%)</b>
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] Exporty 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

<b>Meno</b>	<b>Súčet počtu odhadovaných bodov</b>	<b>Súčet počtu pridelených bodov</b>	<b>Podiel práce (%)</b>
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

## Globálna retrospektíva za 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.

# 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 pozíciách 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) Vyhľadanie 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) Vyhľadanie 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) Vyhľadanie 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) Vyhľadanie 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) Vyhľadanie 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) Vyhľadanie 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 prispôsobili 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 úlohy, o ktorých vieme, že v budúcnosti sa budú musieť implementovať, no v momentálnej dobe nie sú prioritné. Tvorenie nových úloh do backlogu prebieha na stretnutiach tímu, alebo na stretnutiach s vedúcim, kde počas plánovania nového šprintu môže byť niektorá z úloh presunutá do backlogu, ak sa všetci členovia zhodnú, že momentálne existujú problémy s vyššou prioritou.

### **Nástenka**

Na nástenke sú zobrazené úlohy aktuálneho šprintu. Sú rozdelené do troch stĺpcov podľa ich aktuálneho stavu riešenia (to do, in progress a done). Pri každej úlohe 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 danej úlohe 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 posielajú, 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.

## Export evidencie úloh

V nasledujúcej tabuľke sa nachádza evidencia jednotlivých príbehov k dňu ukončenia piateho šprintu (15. 12. 2021). 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	Fuzzy search			To Do	23-11-21 14:40			
Story	Analýza log sash na updaté Elasticsearch z MongoDB a analýza elastic ylm			To Do	22-11-21 11:55			
Story	Zabezpečiť elasticsearch a mongodb kontajner			To Do	21-11-21 19:17			
Story	PDF report z nájdených článkov			To Do	02-11-21 14:50			
Story	Prihlásenie sa na odber nejakého vyhledaného človeka			To Do	02-11-21 14:50			
Story	Označenie článku ako videneho (prečítaného)			To Do	02-11-21 14:49			
Story	Hodnotenie riziku pri nájdených článkoch			To Do	02-11-21 14:47			
Story	Zvýraznenie hľadanych slovčok			To Do	02-11-21 14:45			
Story	Filter na web stránke			To Do	02-11-21 14:32			
Story	Analýza médií, ktoré dostaneme vo výstupoch a aké search queries treba poslať, aby sme mali výsledky ako z Google			To Do	02-11-21 13:25			
Story	Pozrieť sa na zahodené články, kt. neobsahovali paragrafy			To Do	02-11-21 13:23			
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	Zistiť, ako fungujú iné zdroje Google News, prípadne aký mechanizmus používa Google News.		Adam Šľpka	To Do	13-10-21 14:31			
Story	Zistiť, čo spája kriminálne články			To Do	13-10-21 14:13			
Story	Analýza - multilingvistika			To Do	05-10-21 13:37			
Story	Nasu API dat na portal verejných API v rámci propagacie			To Do	05-10-21 13:22			
Story	Technický návrh developerského prostredia	8	David Sliady	IN PROGRESS	02-11-21 13:50		AMS Sprint 4	
Story	User databáza + zbieranie aktivity používateľa		Táňa Poláková	IN PROGRESS	13-10-21 14:29			
Story	Lofi používateľského rozhrania		David Sliady	IN PROGRESS	05-10-21 13:20			
Story	Mongodb unit test	3	Adam Šľpka	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 Sliady	Done	07-12-21 13:36	15-12-21 17:14	AMS Sprint 5	
Story	[RP] Exporty úloh	0	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	Táňa Poláková	Done	07-12-21 13:30	15-12-21 17:09	AMS Sprint 5	
Story	[RP] Metodika docker	1	Adam Šľpka	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 Hlavacka	Done	07-12-21 13:27	08-12-21 11:04	AMS Sprint 5	
Story	[ID] Modul aplikacny server + testovanie	5	David Sliady	Done	07-12-21 13:27	15-12-21 17:14	AMS Sprint 5	
Story	[ID] Modul frontend + testovanie	5	Taha Polakova	Done	07-12-21 13:26	15-12-21 14:57	AMS Sprint 5	
Story	[ID] Modul na indexovanie + testovanie	5	Jakub Hlavacka	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	Rozbehnúť PostgreSQL	3	Jakub Hlavacka	Done	23-11-21 14:50	28-11-21 14:48	AMS Sprint 4	
Story	Komunikacia medzi aplikacnym serverom a API serverom	5	David Sliady	Done	23-11-21 14:46	07-12-21 12:32	AMS Sprint 4	
Story	Komunikacia 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 vysledkov vyhľadavania	13	Taha Polakova	Done	23-11-21 14:43	05-12-21 17:11	AMS Sprint 4	
Story	Komunikacia medzi MongoDB a API serverom	13	Adam Šipka	Done	23-11-21 14:42	07-12-21 12:23	AMS Sprint 4	
Story	Indexovanie stiahnutych dat	8	Jakub Hlavacka	Done	23-11-21 14:41	29-11-21 16:55	AMS Sprint 4	
Story	Dokumentacia k riadeniu projektu		Taha Polakova	Done	23-11-21 9:00	23-11-21 20:39	AMS Sprint 3	
Story	Dokumentacia k inžnierskemu dielu		Jakub Hlavacka	Done	23-11-21 9:00	23-11-21 19:32	AMS Sprint 3	
Story	Rozširenie docker compose o MongoDB a Elasticsearch	5	Jakub Hlavacka	Done	15-11-21 13:16	21-11-21 19:16	AMS Sprint 3	
Story	Deployment scraperu	13	Dominik Horvath	Done	15-11-21 13:08	28-11-21 17:01	AMS Sprint 4	
Story	Implementacia 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	Taha Polakova	Done	09-11-21 14:44	15-11-21 12:50	AMS Sprint 3	
Story	Analýza možnosti 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 Sliady	Done	09-11-21 14:38	23-11-21 14:36	AMS Sprint 3	
Story	Uprava dokumentácie zo stretnutí a sprintov tak, aby mohli ísť na stránku tímu	8	Taha Polakova	Done	09-11-21 14:31	23-11-21 8:51	AMS Sprint 3	
Story	Integrácia MongoDB a Elasticsearch	8	Jakub Hlavacka	Done	09-11-21 14:22	16-11-21 12:54	AMS Sprint 3	
Story	Integrácia scraperu a vylepšeného parseru	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	Vyhľadanie RSS dvojičata k HTML stránke - všeobecný postup		Dominik Horvath	Done	02-11-21 14:08	09-11-21 9:45	AMS Sprint 2	
Story	Napínanie MongoDB priamo pri scrapovaní		Adam Šipka	Done	02-11-21 13:44	14-11-21 14:06	AMS Sprint 2	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 2	
Story	Pamätová zložitosť elastic search na vzorke dát		Jakub Hlavacka	Done	01-11-21 12:15	12-11-21 17:09	AMS Sprint 2	AMS Sprint 3
Story	Parseovanie tagov pomocou lxml, readability, traflatura		Taha Polakova	Done	30-10-21 11:41	09-11-21 13:12	AMS Sprint 2	

Story	Zavesenie klienta	13	David Sliady	Done	27-10-21 14:45	08-11-21 0:51	AMS Sprint 2	
Story	Rozšíriť docker compose	13	Dominik Horvath	Done	27-10-21 14:43	31-10-21 20:22	AMS Sprint 2	
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 2	
Story	Request URL s odpoveďou inou ako 200 uložiť do súboru	3	Adam Šípka	Done	27-10-21 14:40	01-11-21 11:35	AMS Sprint 2	
Story	Lokálna MongoDB so získanými dátami	8	Adam Šípka	Done	27-10-21 14:35	01-11-21 11:35	AMS Sprint 2	
Story	Analýza bezstratovej kompresie textu	8	Jakub Hlavacka	Done	27-10-21 14:25	01-11-21 12:25	AMS Sprint 2	
Story	Parsovanie tagov pomocou regex	8	Táňa Poláková	Done	27-10-21 14:19	02-11-21 13:49	AMS Sprint 2	
Story	Parsovanie tagov pomocou CSS Selector	5	Jakub Müller	Done	27-10-21 14:17	08-11-21 19:48	AMS Sprint 2	
Story	Stránka tímu		Jakub Müller	Done	18-10-21 15:03	01-11-21 11:37	AMS Sprint 1	
Story	Otvorenie portov na serveri		David Sliady	Done	18-10-21 15:01	18-10-21 17:05	AMS Sprint 1	
Story	Vyriešiť, aby flitkar nemohol 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	AMS Sprint 1	
Story	Preiskum databáz		Jakub Müller	Done	13-10-21 14:26	18-10-21 14:28	AMS Sprint 1	
Story	Vytvoriť github projekt		Jakub Müller	Done	13-10-21 14:22	18-10-21 12:09	AMS Sprint 1	
Story	Príháška na TP CUP			Done	13-10-21 14:13	02-11-21 12:50	AMS Sprint 1	
Story	Set up elastic search v docker container		Jakub Hlavacka	Done	13-10-21 14:12	16-10-21 16:45	AMS Sprint 1	
Story	Získanie dát pre prototyp		Dominik Horvath	Done	13-10-21 14:09	26-10-21 13:29	AMS Sprint 1	
Story	Vytvoriť Slack server		Táňa Poláková	Done	05-10-21 13:49	11-10-21 12:08		
Story	Analýza parametrov Google News		Adam Šípka	Done	05-10-21 13:43	13-10-21 14:26		
Story	Realny scrapping - vytvorenie vzorky		Dominik Horvath	Done	05-10-21 13:27	10-10-21 18:14		
Story	Zísť, ktoré miestnosti su na karticky		David Sliady	Done	05-10-21 13:24	18-10-21 16:12		
Story	Revizia požiadaviek		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	Prípady použitia		Dominik Horvath	Done	01-10-21 14:23	05-10-21 12:08		
Story	Prieskum získavania dát, práca s knižnicou		Adam Šípka	Done	01-10-21 14:22	05-10-21 13:24		
Story	Specifikácia požiadaviek		Jakub Hlavacka	Done	01-10-21 14:22	03-10-21 15:17		
Story	Vysoka architektura		David Sliady	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		